



Field Guide for Data Quality Management

Field Guide for Data Quality Management

Monitoring, Evaluation, Results and Learning Series Publications

Module 2

Pact, Inc.

Washington, DC

November 2014

Copyright © 2014 by Pact, Inc.



This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License.
To view a copy of this license, visit
<http://creativecommons.org/licenses/by-sa/4.0/>
or send a letter to Creative Commons, 444 Castro Street, Suite 900,
Mountain View, California, 94041, USA.

Contents

Contents	1
Foreword.....	2
How to Use This Manual.....	2
Quick Reference Guide to Essential Data Quality Management Concepts.....	3
Acronyms	4
Chapter 1: Introduction to Data Quality Management Concepts	5
Chapter 2: Data Quality.....	8
EXERCISES.....	18
#1: Data Quality Issues.....	18
#2: Data Quality Criteria.....	18
Chapter 3: Data Management.....	19
EXERCISES.....	29
#3: Challenges and Solutions to Data Quality Management	29
#4: Data Flow Mapping Exercise	31
Chapter 4: Data Quality Assessments and Audits	32
EXERCISE.....	37
#5: Identifying Key Data Quality Assessment Steps for Your Program.....	37
Appendix 1: How to Use the Routine Data Quality Assessment Tool.....	38
Appendix 2: Data Quality Management Plan TEMPLATE	49
References	57

Foreword

Data quality is a cornerstone of accountability in program reporting. In the international development sector, although we are often focused on reporting, ensuring the quality of the data that we report is critical for our partners, our donors, and our beneficiaries. In addition, Data Quality Management Plans and Routine Data Quality Assessments are both important elements of Pact's Results and Measurement Standards. The intent of this manual is to provide guidance on how to ensure excellent data quality in all our programming. A slide set accompanying the module provides an opportunity to engage in practical exercises to test the skills outlined in this text.

How to Use This Manual

Chapters 1 through 4 of this manual will provide Pact Staff with a solid understanding of how to assess data quality and how to best conduct data management for data quality. The shaded boxes at the beginning of each chapter outline the key learning concepts and the exercises at the end of each chapter will help you begin formulating aspects of your project's Data Quality Management Plan. In the annexes you will find:

- Instructions on how to use the Excel-based Routine Data Quality Assessment (RDQA) Tool—to use when conducting RDQAs of your own data and M&E systems, as well as your partners' data and M&E systems;
- A Data Quality Management (DQM) Plan template to customize to your own program.

This manual was updated and revised in 2014 to reflect field experience with routine data quality assessments and Pact's own internal expertise in improving data quality. The updated manual was revised by Lauren Serpe, Alison Koler, Reid Porter, Rachel Beck, and Jade Lamb. Copyediting was done by Karen Cure. With the exception of a new RDQA Tool, much of the original manual's content remains, and I would like to thank Lynn McCoy, Rita Sonko, Hannah Kamau, Jacqueline Ndirangu, Titus Syengo, and Ana Coghlan for their contributions.

Kerry Bruce

Senior Director, Global Health and Measurement

rm@pactworld.org

Quick Reference Guide

to Essential Data Quality Management Concepts

Although the coming chapters will cover many of the following terms in detail, they refer to common data quality management concepts that are helpful to be familiar with from the beginning.

Audit trail: A collection of documents and notes that help clarify exactly how data results were derived.

Data quality assessment (DQA) (or Routine data quality assessment-RDQA): A procedure that provides an organization with the means to determine the status of data quality at any given time and the opportunity to develop and implement strategies to address any gaps.

Data quality management: The management of the data system, comprising six key stages: data source, data collection, data collation, data analysis, data reporting, and data usage.

Data quality: The worth or accuracy of the information collected. The term emphasizes the importance of ensuring that the process of data capturing, verifying, and analysis of data is executed to a high standard, such that it would meet the requirements of an internal or external DQA or audit.

Data quality audit: An official, rigorous inspection (often by a funding agency) of program data to determine its reliability, validity, and overall level of excellence.

Face validity: The existence of a solid, logical relation between the activity or program *and* what is being measured.

Measurement validity: The accuracy of data measurement, arising from essential qualities of data measurement tools and procedures—that is, that they are well designed, defensible, and limit the potential for errors.

Reliability: The extent to which data collection processes are stable and consistent over time—usually as a result of *internal quality controls* in place and transparency of data procedures.

Standard operating procedure (SOP): A written document or instruction detailing relevant steps and activities of a process or procedure. An SOP provides employees with a reference to common practices, activities, or tasks.

Transcription validity: Soundness of data entry and collation procedures, ensuring that data are entered (transcribed) and tallied correctly.

Validity: The extent to which a measure actually represents what it is intended to measure. Three types of validity are important to know in data quality management: face validity, measurement validity, and transcription validity.

Acronyms

DQA	data quality assessment
DQM	data quality management
DQSO	data quality strengthening objective
IRB	institutional review board
M&E	monitoring and evaluation
MERL	monitoring, evaluation, research, and learning
MIS	management information system
NGO	nongovernmental organization
OCA	organizational capacity assessment
OS	other stakeholders
OVC	orphans and vulnerable children
PEPFAR	US President's Emergency Plan for AIDS Relief
RDQA	routine data quality assessment
RDQM	routine data quality management
SOP	standard operating procedure
USAID	US Agency for International Development
USG	United States Government
VR IPT-CCE	validity, reliability, integrity, precision, timeliness, completeness, confidentiality and ethics
WHO	World Health Organization

Chapter 1: Introduction to Data Quality Management Concepts

In this chapter, readers will learn the key concepts to be covered in the rest of the manual:

- Definition of data quality
- Definition of data quality management
- Definition of Routine Data Quality Assessments
- Definition and elements of a Data Quality Management Plan

What Is Data Quality?

Data quality refers to the accuracy or worth of the information collected and emphasizes the high standards required of data capture, verification, and analysis, such that they would meet the requirements of an internal or external data quality audit.

Data quality grows out of an organization's commitment to the accuracy of data and to ensuring data utility for program decision making and for accountability to donors. Ensuring high-quality data is important, whether the purpose of your monitoring and evaluation is to use data for decision making, to improve organizational programming and learning, or to accurately report your work to your beneficiaries, board, donors, or staff.

To ensure accuracy in your data, it is not enough that you select the best indicators and write high-quality protocols. If you do not use these tools properly, data can still be of poor quality. Reporting standards for both quality and timeliness must be respected.

Every organization needs to develop and document its methods for checking data quality.

The process of checking data quality is often referred to as a *routine data quality assessment* (RDQA), or sometimes as a *data quality audit*. RDQAs help identify where data quality is poor and point to potential solutions. Issues and risks relating to data quality need to be thought through and documented to ensure that quality standards are developed and maintained.

Commonly Used Criteria for Assessing Data Quality

Data quality is most commonly assessed in terms of five key criteria: validity, reliability, integrity, precision, and timeliness. Pact also recommends data quality is assessed in relation to completeness, confidentiality and ethics. Throughout the text, these concepts will be referred to as **VR IPT-CCE**.

What is Data Management?

Managing data means thinking about how data cycle through the organization: controlling how the data are collected and how the raw data are assembled and analyzed; determining the most appropriate presentation formats for the data; and ensuring data use by decision makers. Six key stages make up this *data management cycle*: data source, data collection, data collation, data analysis, data reporting, and data usage.

What is a Routine Data Quality Assessment (RDQA)?

The RDQA is an essential procedure that allows an organization or donor to determine data quality at any given time and the opportunity to develop and implement strategies to address and prioritize gaps. The process consists of asking pointed questions on data quality and data management processes and researching the answers. By asking these questions, the organization can determine a data set's potential for error and therefore understand how confident the staff can be with the results and in using the data to evaluate the program and make management decisions.

Routine data quality assessments (RDQAs) are conducted by the project and have more room for flexibility, whereas DQAs conducted by an external party, such as a funding organization, will follow those requirements. RDQAs have three primary components:

1. *Data Management Review*: Are data management systems and procedures in place adequate to ensure data quality?
2. *Data Verification/Indicator Assessments*: Are the data being collected accurate?
3. *Developing a Data Quality Action Plan*: If there are problems with data verification or data management, how should the organization proceed? What areas should be prioritized for improvement? Who should be responsible for following through on these actions?

What Is a Data Quality Management Plan?

A data quality management plan brings together how to manage data for data and how to assess data quality through assessments. **A data quality management plan is a document that explains your approach to maintaining data quality standards.**

DQM plans are an important component of an M&E system. They are where an organization outlines what data it will collect and how it will ensure quality data, manage the data, and archive the data. The DQM Plan can be incorporated into the project's MERL plan or it can be a stand-alone document. Please view the latest MERL Standards for further guidance on DQM Plans.¹

Format of a Data Quality Management (DQM) Plan

A DQM plan consists of an introduction, data management process description, description of routine monitoring system, and a section of reports. A DQM plan should pull together all

¹ *Pact Quality Standards for Results and Measurement*. <https://pactworld.jiveon.com/groups/results-and-measurement>

of the elements of data quality. The exercises in Chapters 2–4 are designed to be useful in formulating the elements of your DQM Plan, and the template in **Appendix 2** can be used as the starting point for your project’s document. When complete, this document should be part of the project’s PMP/MERL Plan, should be accessible to staff and volunteers, and they should understand it well.

Introduction

The introduction gives a project overview and discusses the DQM plan purpose. It may also include a table of key indicators and an overview of the stakeholders and other personnel who will be involved in data flow.

Data Management Process Description

This section includes the data flow map and the data use plan.

Description of Routine Monitoring Systems

This section covers data quality concerns, a list documenting the frequency of site visits, and a reference to the RDQA tool to be used during site visits.

Reports and Action Plans

This section outlines the frequency and format of data quality-related reports. In particular, this section may offer a template for RDQA action plans, describe how often they should be issued, and provide steps for creating and implementing DQA action plans.

What Are Standard Operating Procedures (SOP)?

An SOP is written documentation detailing all relevant steps and activities of a process or procedure. An SOP provides employees with clear written reference to common practices, activities, and tasks and ensures consistent practices in data quality management. You can use SOPs to mentor staff, partners, and volunteers to follow routinely, with the goal of keeping data quality consistently high. For Pact’s purposes, “Best Practice SOPs” will usually be integrated throughout a project’s Performance Monitoring Plan document, elaborating how the project ensures data quality at each step of the data management process. Generally, it is most useful when you spell out these best practices per indicator or data collection process. Chapters 2 and 3 of this manual will list some of these “Best Practice SOPs” that can be integrated into other parts of a larger PMP. Sometimes, projects will need to develop separate SOPs to detail out specific procedures beyond what is described in the PMP. This may be necessary if the donor requires these documents, or if the description in the PMP is not sufficient. If your project requires separate SOPs, you can reach out to your M&E advisor for guidance around developing those documents.

Chapter 2: Data Quality

In this chapter, readers will learn:

- Eight key criteria for assessing the quality of data.

In building data quality management systems in this chapter, readers will:

- Identify your organization's key data quality issues (Exercise 1).
- Assess understanding of data quality criteria (Exercise 2).

Validity

The first data quality criterion is *validity*. For a data set to be valid, we need to ensure that data adequately represent performance. The key question on validity is whether the data actually represent what they are supposed to represent. For example, if 52% of community workers have been trained in psychosocial support, the assumptions are that:

- **We calculated the correct percentage:** 52 %—*not 41% or 60%*;
- **We counted the correct beneficiaries, correct training, and & only those completing training:** We counted targeted community workers, all those our program intended to work with *and not including their neighbors, our staff, or government personnel*; and those who have actually completed—*not just started or been invited to*—a training in psychosocial support—*not in a different subject*;
- **We count what we intended to count:** what we intended to measure was the number of people trained—*rather than, for instance, the number of people actually providing psychosocial support*.

If all this holds true, then the data are valid and adequately represent performance. If the data are a sample of the population, rather than a census or a specific case study, there needs to be certainty that there are no significant measurement or representation errors. To ensure that the sample is representative, sampling methods must be accurate; response rates must be high enough; the population sampled must be appropriate; and anyone collecting data must be appropriately trained.

Validity can be assessed by making sure there is adequate *face validity*, *measurement validity*, and *transcription validity*.

Face Validity

Face validity refers to a solid, logical relation between the activity or program and what is being measured. This means that the “right” indicator has been selected and that that indicator measures what it is intended to measure. For example, if we want to know *x*, does it make sense that we are measuring *y*? Face validity has to do with making sure the indicators selected are both *direct measures* (i.e., the indicators are closely aligned with what we actually want to know) and *relevant* (i.e., the indicators are capable of providing evidence to prove or disprove whether the changes measured are caused by the organization's

activities, or at least to some extent attributable to them and are not collected simply for the sake of collecting data).

Measurement Validity

For data to have *measurement validity*, data measurement tools and procedures must have been well designed and defensible and must limit the potential for errors. The following are some examples of measurement validity errors to watch out for:

You Could Have *Data Validity* Issues If You Answer “Yes” to These Questions

- Did respondents have trouble understanding the questions that were asked of them?
- Are data incomplete and illegible?
- Did respondents feel pressured to answer correctly?
- Were data altered in transcription?

Sampling or Representation Errors: *Sampling errors* arise as a result of drawing a sample that does not represent the population served. For instance, the data collected from young people will not accurately describe older generations. To avoid this, the sampling frame (i.e., list of units in the target population from which the sample is selected) must be up to date, comprehensive, and mutually exclusive for separate geographic areas. Sufficiently high response rates and additional follow up with non-respondents are necessary to ensure that all groups are adequately represented.

Nonsampling Errors: Mainly associated with data collection and processing procedures, from such issues as interviewer bias and self-presentation bias—often arise as a result of misleading definitions and concepts, unsatisfactory questionnaires, incomplete coverage of sample units, or defective methods of data collection, tabulation, or coding.

Memory errors (or *recall bias*), a subset of nonsampling errors, occur when items in an inquiry relate to events that happened in the past and the respondents either fail to remember them or place them in the wrong time periods. Memory errors may be a function of the time between the inquiry and the time when an event occurred.

Transcription Validity

For data to have *transcription validity*, the entry and collation procedures must have been sound, with limited potential for error; steps must have been taken to limit potential for transcription errors; and the data must have been transcribed, entered, and tallied correctly. This can be handled by many programs and databases that have data enter checks and logical error checks. It can also be handled by conducting double data entry on all or a sample of your data to check for errors. Data processors should be told how to handle errors—how to document and communicate transcription errors, track them back to the source, and develop a plan to reduce the error rate. Many organizations have been misled simply because data were not input into the database properly.

Reliability

The second data quality criterion is *reliability*. For a data set to be reliable, data collection processes must be stable and *consistent* over time, with *reliable internal quality controls* in place and data procedures handled in a *transparent* manner.

Consistency: For data to be reliable, they must be *consistent*:

- Consistent data collection *processes* are used from year to year, location to location, data source to data source.
- The same *instrument* is used to collect data from year to year, location to location.
- If data come from different sources, the instruments must be similar enough to make the data collected comparable.
- The same *sampling method* is used from year to year, location to location, data source to data source.
- When raw data need to be manipulated to produce the data required for the indicator, the same formulae are applied consistently (and correctly) from year to year, site to site, data source to data source (for example, when data from multiple sources need to be aggregated).
- Procedures are in place to deal with missing data, and final numbers are double-checked for accuracy.

Internal Quality Controls: For data to be reliable, *internal quality controls* must be in place:

- Procedures must ensure that data are free of significant error and that bias has not been introduced.
- Procedures must dictate periodic review of data collection, maintenance, and processing.
- Procedures must outline periodic data quality assessments.

You Could Have Data Reliability Issues If You Answer “Yes” to These Questions

- Did you use different tools to collect the same data?
- Is recording inconsistent from one staff member to another?
- Were instances of wrong or missing data not reported to the organization?

Transparency: For data to be reliable, *transparency* can be ensured through:

- Data collection, cleaning, analysis, reporting, and quality assessment procedures must be documented in writing.
- Data problems at each level must be reported to the next level.
- Data quality problems must be clearly described in final reports.

Integrity

The third data quality criterion is *integrity*. For a data set to have integrity, the data must be accurate and free of error introduced by either human or technological means, either willfully or unconsciously. Specifically, we need to make sure no manipulation or bias has been introduced into the data set.

Unconscious Integrity Issues: These can arise when organizations offer incentives to encourage data collection timeliness, resulting in rushed or sloppy data collection or entry.

- Positive incentive example: “30 minutes of free phone time will be given to all volunteers who submit their data on time.”
- Negative incentive example: “Employees will have all data on my desk by Friday 5 pm or else!”

You Could Have *Integrity* Issues if You Answer “Yes” to These Questions

- Has anyone tried to bias or influence the outcomes the data present?
- Are there unreasonable time pressures to produce your data?

Willful Integrity Issues: These occur when a person or organization purposely provides false data (e.g., fills out a survey without talking to a respondent), as when data collection staff or volunteers are unsupervised, are poorly trained, or feel pressured to submit their material, either in terms of time or a political agenda. Data collectors always need to be told that it is better not to submit any data than to submit false data. The use of mobile data collection can significantly decrease the likelihood of falsified data due to GIS-enabled settings—showing the data supervisor exactly where enumerators have collected each form.

Key Questions to Probe for Manipulation or Bias Affecting Data Integrity

- Are objectivity and independence protected for in key data collection, management, and assessment procedures?
- Are mechanisms in place to reduce the possibility that data could be manipulated for political or personal reasons?
- Are you confident in the credibility of data from a secondary source?
- Has the data ever been independently reviewed by a person outside the data collection effort and/or of the organization collecting the data?
- Is there any *interview bias*—distortion of responses caused by interviewers’ different styles, personalities, or presentation of particular questions?
- Is there any *self-presentation bias*? Also known as social-desirability bias, this is when people modify their thoughts so as to give the answer the interviewer will think is “right.”

Precision

The fourth data quality criterion is *precision*. According to USAID TIPS 12², “Precise data have a sufficient level of detail to present a fair picture of performance and enable management decision making.” For a data set to be “precise,” relevant data should also be collected by the designated disaggregation characteristics, such as sex, age, and geographic location. If data is based on a sample, the *margin of error* for the data must also be acceptable. Margin of error, expressed in a percent, is a measure of the data’s uncertainty or our level of confidence in the data. For example, an acceptable margin of error may be $\pm 5\%$ at a 95% confidence interval. Many times for our purposes, margin of error can refer to

² http://pdf.usaid.gov/pdf_docs/pnadw112.pdf

targets set for indicator data and how confident we can be that the change we wanted to see is valid. When referring to margin of error in sampling³, the larger the margin of error, the less confidence one should have that the data reported are close to the “true” figures; that is, the figures for the entire population. How can you calculate what’s acceptable?

You Could Have Data Precision Issues If You Answer “Yes” to These Questions

- Did only one staff member aggregate data, without another staff member’s review?
- Was data only collected at the aggregate level, despite promising to collect disaggregated data?
- Is the margin of error larger than the change being measured?
- Was the margin of error not reported?

- The margin of error should be less than the expected change being measured. If the data show that HIV incidence among the sample is 1.5%, the margin of error should also be less than 1.5%.
- Another example from USAID Tips 12: The number of politically active non-governmental organizations (NGOs) is 900. Preliminary data shows that after a few years this had grown to 30,000 NGOs. In this case, a 10 percent measurement error (+/- 3,000 NGOs) would be larger than the original number. It may not be important to know precisely whether there were 29,999 or 30,001 NGOs. A less precise level of detail is still sufficient to be confident in the magnitude of change. However, if the growth in politically active NGOs after a few years was only up to 1,000 a 10 percent measurement error (+/- 100) would not be acceptable because it would not be able to tell us if anything had changed at all.
- Take into account the likely management decisions to be affected. Consider the consequences of the program or policy decisions based on the data—how narrow does the range need to be for the data to be useful?
- Targets should be set for an acceptable margin of error before the data are analyzed.
- The margin of error should be reported along with the data. This enables the people using the data to make their own assessment of the acceptability of the data’s precision level.
- If the data is based on a sample, consider whether the cost of increasing the precision of the data by increasing the sample size is worth the increased accuracy possible. For more information on calculating margin of error, there are numerous resources and calculators online.⁴

Timeliness

A fifth data quality criterion is *timeliness*. To be considered “timely,” the data must be collected *frequently* enough and must be *current*.

Key Questions to Ask about the Frequency of Data Collection

- Are data collected frequently enough to inform program management decisions?

³ http://www.statistics.com/glossary&term_id=393

⁴ <http://americanresearchgroup.com/moe.html>

- Is a schedule of data collection, collation, analysis, and reporting in place that meets program management needs?

Key Questions to Ask about Data

Currency

- Are the data reported in the most current time frame practicably available?
- Are the data from within the policy period of interest?
- Are the data aggregated, analyzed and reported as soon as possible after collection?
- Is the date of data collection clearly identified in reports?

**You Could Have *Timeliness* Issues
If You Answer “Yes” to These Questions**

- Is it taking too long to get data in the database, or to develop a data report?
- Are decisions being made without data because it’s never available on time?
- Are the data so old when reported that they have lost much of their relevance and value?

Additional Criteria: Completeness, Confidentiality and Ethics

Data quality is most commonly assessed in terms of checking data against the five key criteria described above. Although the US President’s Emergency Plan for AIDS Relief (PEPFAR), via the US Agency for International Development (USAID), as well as most European donors accept validity, reliability, integrity, precision, and timeliness as the foundation for understanding data quality, additional factors are sometimes considered as well: completeness, confidentiality, and ethics, which will be covered below.

Completeness

How complete is the data set? *Completeness* refers to the degree to which all the necessary steps in data collection, data entry, data cleaning, and data analysis have been carried through. In addition, no data are missing and no responses are incomplete, uncollected, or, because of other data quality issues, unusable.

Confidentiality

The Global Fund to Fight AIDS, Tuberculosis and Malaria—an international financing organization—uses a seventh criterion in addition to the six discussed above: *confidentiality*. In assessing your data against this standard, consider whether you and your team have in place and utilize processes and systems ensuring that data are collected, stored, and reported in a way that protects respondents’ privacy. Privacy is especially important when you are collecting personally identifiable, sensitive, or geographic data—as would be the situation, for example, if you were pinpointing on a map the homes of all those living with HIV in a specified area.

Ethics

Ethics are a system of moral ideas and rules about our conduct that reflect international standards and the values of the culture we work in and of the communities we serve. Data

ethics are the rules or standards governing the conduct of a person collecting, collating, reporting on, or utilizing data, and represent our standard of what's "right."

The data management process involves a series of decisions made by the organization to ensure that all work with the data conforms to the highest ethical standards. Informed consent, privacy, confidentiality, and integrity are basic ethical tenets of any data collection effort. You and your organization are responsible for ensuring the ethical treatment of participants.

Data ethics are an important part of having a quality data set. It's also important to consider:

- Are procedures for ethical review of the data collection protocols in place?
- Have these procedures been reviewed by a community advisory board, if appropriate?
- Is an independent, culturally appropriate ethical review board available to evaluate planned data collection, collation, reporting, and utilization protocols?
- Has the organization established and followed standards governing the conduct of the people collecting, collating, reporting, and utilizing data?
- Have processes for obtaining informed consent been put into place, both for adults and for children?

Common Types of Ethical Issues

Informed Consent

Informed consent is a basic human right: an agreement to participate in a data collection effort, or not, that is based on an appreciation and understanding of the facts and implications of the decision. When an individual or community has given informed consent, it means that they feel they have enough information about the data collection effort to enable them to make knowledgeable decisions about whether or not to participate.

Participants need not only to be aware of the relevant facts, but also must be considered to have reasonable judgment and no impairment of their reasoning faculties. Participants must also act on their own behalf. For instance, children cannot give informed consent to participate in an evaluation exercise; rather, their parents must join them in agreement regarding their participation in the data collection activity. Impairments might include age, illness, intoxication, insufficient sleep, mentally retardation, mental illness, or other health problems.

In cases where an individual does not possess all the relevant facts regarding the data collection process, serious unforeseen harm may arise. The harm may take the form of stress caused by the burdens of participation in the data collection effort, loss of social status, loss of self-esteem, psychological injury, or other side effects. Even in cases where consent from participants is provided, an organization still has an obligation to protect the respondent as far as possible against potentially harmful effects of participating.

Informed consent is a process that involves:

- The organization providing all relevant information to the potential participant.
- The potential participant understanding the information provided.

- The potential participant willingly agreeing to participate.

Do You Really Have Informed Consent?

- A person may verbally agree to something from fear, perceived social pressure, or psychological difficulty in asserting her/his true feelings. The person requesting the action may honestly be unaware of this. *Consent is expressed, but is not actually internally provided.*
- A person may express an understanding of the implications of some action while giving consent, but in fact has not fully appreciated the possible consequences and may later deny the validity of her or his consent for this reason. *Understanding needed for informed consent is stated to be present but is in fact (through lack of understanding) not present.*
- A person below the age of consent may agree to participate and be fully aware of all the consequences, but the consent is invalid because, due to age, the individual is deemed to have immature judgment and therefore be incapable of giving truly informed consent.

Protection of Privacy and Confidentiality

Data collection often requires people to reveal aspects of their behavior or attitudes to the researcher. And although study subjects may be willing to share such information with researchers, they may not be willing to have that information more broadly known. During the informed consent process, participants need to be made aware of the degree of privacy they can expect.

Privacy refers to the right of individuals to limit access by others to aspects of their person (i.e., as to themselves or things that belong to them). “Belongings” in this context might include, for example, their thoughts, ideas, attitudes, beliefs, and identifying information such as their name, geographic location, or the like.

Confidentiality involves the researcher’s protecting the participant’s private information. Even if the researcher knows where or who specific data came from, the information will remain held closely by the researcher.

Often a researcher may promise *anonymity* to a participant. That means that a person’s name and or other identifiers (e.g., their address) will not be attached to the data or ever presented with the data.

Anonymity alone by no means guarantees confidentiality. Remember that a particular configuration of attributes (e.g., membership in an organization combined with number of children, or a particular success story) can, like a fingerprint, frequently identify an individual beyond a reasonable doubt. Organizations need to counteract opportunities for others to infer identities from their data.

Misrepresentation or Falsification of Data

Organizations depend on the confidence of the communities, their partners, and donors. Great care needs to be taken to protect an organization's reputation, and data should never be presented in a way that exaggerates that data's accuracy or explanatory power. Having data that are incomplete, anecdotal, or based on case studies and suggesting that the data are a completed data set or representative sample is *misrepresentation*. Failure to note data limitations or to cite data sources is also misrepresentation. Adding missing data during collation, making up data, and changing responses are falsification of data and ethically incorrect.

Organizational Responsibility for Data Ethics

Data management involves a series of decisions made by an organization to ensure ethical data collection and thereby the ethical treatment of participants.

To help ensure ethical data collection, it's important for organizations to have routines that ensure best practices are observed and at the same time to mentor partners and grantees to do the same. Here are some examples:

- Carefully evaluate ethical acceptability of all data collection efforts. Determine whether participants are at any risk.
- Provide a description of the study in advance, including: the purpose of the research, expected benefits, methods (tasks to be performed), any effects, risks, or inconveniences from the procedures, participant rights, and any alternative procedures.
- Ensure informed consent. Always provide participants (both individuals and communities) with enough information about the data collection effort that they can make an informed decision about whether they want to participate or not.
- Always obtain a written consent statement (or verbal consent, if applicable). Consider the rights of children and their capacity to provide informed consent—obtain permission from both the children and their parents or guardians.
- Protect privacy. Always make participants aware of how much privacy they can expect, be it confidentiality or anonymity. Do not discuss an individual's answers, and be sure neither staff nor community volunteers observe the same protocol.
- Use “blind data.” Do not put names on the data in files or reports. Use codes if possible.
- Never misrepresent or falsify data. This is a serious offense and could be grounds for a review of staffing and grantee funding if the following occur:
 - Inappropriate scale-up: Presenting incomplete, anecdotal, or case study data as a completed data set or representative of a larger population.
 - Failure to note the results are due in whole or part to a partner.
 - Adding missing data during collation; making up data; or changing responses.
 - Data limitations not made clear during reporting.
- Ensure that staff and grantees are aware that all participation in data collection efforts is voluntary. No one may be refused services because he or she does not want to provide information. All subjects have the right to refuse to participate in data collection efforts. Never coerce people into participating.

- Have a signed code of conduct for all individuals assisting in the data collection, collation, analysis, and reporting effort.
- Empower and include communities. Discuss with communities and participants how you plan to carry out the collection effort and why you are doing the research. Discuss with communities and participants how you will use the data. Know and follow cultural norms and practices as appropriate.
- When presenting results, explain the limitations of the findings. Provide results and feedback to the participants and the communities.
- Remember that ethical codes apply to use of pictures, success stories, and GIS data.
- Know and follow any relevant research ethics policies, including those of established by your organization and by relevant national or international bodies.

Ethical Guidelines

The US government has a guideline called the “Common Rule” to which USAID is a signatory. This guidance, based on the Belmont Report provides an overview of the regulations that all US funded programs much follow when conducting research with human subjects. See the HHS.gov Web site for a link to the Belmont Report and further information at:

<http://www.hhs.gov/ohrp/humansubjects/commonrule/index.html>

EXERCISES

#1: Data Quality Issues

Take a few moments to think about data quality issues in your own program. Gather into small groups, and discuss for 30-45 minutes, then present your work to the larger group.

What issues do you face in collecting and reporting good quality data? What are some of the common problems with your data?

What data quality issues are special to the technical area of your program (e.g., orphans and vulnerable children, advocacy, peace building, microfinance)?

Capture on a flip chart at least 10 key data quality challenges you and your peers face. What solutions could you put into place to address those issues?

#2: Data Quality Criteria

This exercise is a continuation of exercise 1. Using the lists of data quality challenges that you developed in the first exercise, examine each challenge and determine which of the eight main criteria apply to each challenge. Once you have decided, please label each challenge with the first letter of the criteria that apply (V,R,I,P,T Com, Con, or E). More than one criterion may apply to each challenge. Example:

Inconsistent use of data collection tools	R
Level of understanding of data collectors	
Low level of education	I
Do not understand importance of collecting data	R V P
Do not understand indicator definitions	R
Delayed reporting	T
No reporting timeline	T
No source documents	I P
Incomplete data	P R
Double counting or under reporting	P
No verification process	R I
Staff turnover	I R P
Collection personally identifiable information	Con, E
No proper documentation or data storage	P R

Chapter 3: Data Management

In this chapter, readers will learn:

- The six key data management stages.
- Best practices for data management at each process level.

In building a data quality management plan, readers will:

- Identify solutions to problems within each data management stage and begin thinking about best practices you can implement (Exercise 3)
- Map the flow of data in their organizations (Exercise 4).

Data Management

To ensure data quality, an organization needs to comprehensively and effectively manage its data. The management process involves analyzing factors that might positively or negatively influence data quality at each stage of data cycle, so as to protect the data from corruption at any point and to maintain the data quality at a high level.

A. Six Stages of the Data Management Cycle

Managing data means thinking about how data cycle through the organization: controlling how the data are collected and how the raw data are assembled and analyzed; determining the most appropriate presentation formats for the data; and ensuring data use by decision makers. Six key stages make up this *data management cycle*:

1. Data source
2. Data collection
3. Data collation
4. Data analysis
5. Data reporting
6. Data usage

1. Data Source

Data can come from different sources: a person's opinion expressed for a survey, a signature on a sign-in sheet, a patient or student file, or a national study or report completed by another organization. To gauge the *quality* of a given data set, you need to understand exactly where the data originated and the strengths and weakness of the data at that source.

Data sources are classified as primary, secondary, or tertiary.

Primary Data Sources

Primary data are usually the first source of data available. It is firsthand information—original and created by individuals who participated in the events being studied, usually either during or shortly after the time period of interest.

For example, if an organization wanted to count the number of people trained in a training program, the sign-in sheets would comprise the primary data; if an organization wanted to know the number of visits a volunteer made to an orphan, the primary data source would be the record of services in the volunteer's notebook; if a hospital wanted to know where its patients were coming from, the primary data source could be their addresses in the institutional files.

Primary data present information in its original form, not interpreted by others.

Lowest Risk to Data Quality: Primary data sources present the least risk to potential errors as the data are raw, un-manipulated, and firsthand. Wherever possible, it's preferable to tally and use reported numbers from primary data sources rather than from secondary sources. Failing that, secondary data should be routinely checked.

However, primary data still present quality risks if not routinely assessed against standard data quality operating procedures.

Secondary Data Sources

A source is considered secondary when the data originated somewhere else and were transferred into another format and reported by someone else. An example is the Excel spreadsheet created by a consultant trainer who typed in the list of trainee names off the sign-in sheet; the organization who hired the consultant is not using primary data—data in their original form or collected directly by the organization—but can use the data to monitor or evaluate the training program.

Commonly used secondary data sets—all publicly available—include those from the World Health Organization (WHO), the World Bank, MEASURE DHS, and the U.S. Centers for Disease Control and Prevention.

Secondary sources can simplify the data collection process if the data quality aligns with the five measures of data excellence. Already collated, cleaned, and ready for analysis, secondary data can significantly decrease the cost and difficulty of data collection. However, high-quality secondary data sources relevant to program monitoring and evaluation are sometimes difficult to find. And even when the data sets themselves are available, the lack of knowledge as to the process behind their collection and collation may make it impossible for an organization to reliably assess their quality.

Secondary data is useful to triangulate your own data. *Triangulation* refers to using known data (such as district service data) to estimate the possible achievement for an indicator. If only 500 clients were served in a district during a set time period, is it likely that your program served 10,000 in the same period?

Tertiary Data Sources

Tertiary data are one or more steps removed from the secondary data. This information often indexes citations to secondary sources (and sometimes primary sources), and shows you how to use them. Tertiary data are often materials in which the information from secondary sources has been digested into a convenient, easy-to-read form, such as abstracts, articles, and manuals.

The following are some best practices projects can follow to ensure data quality at the **Source**. These best practices can also be adopted or tailored for SOPs throughout your PMP:

Best Practices for Ensuring Data Quality at the Source

- Include data providers, community stakeholders, and data processors in the design and testing of approaches and tools.
- Choose indicators that measure the outcomes of interest as directly as possible, staying mindful of the needs of management. Define and clarify the indicators.
- Develop guidance around engaging with stakeholders. For example, before data collection begins specify how the team should confer with communities about the process.
- Develop guidelines for the design of data collection instruments. For example, for questions that require recall, provide answers with specific reference points (e.g., “a month ago” or “a year ago”) rather than a range (e.g., “three to six months”) or a subjective timing (e.g., “recently”). Conduct pretests or pilots to check clarity and respondents’ recall.
- Develop and document instructions for data collectors and computer procedures, and include them on forms. Be explicit about guidance on training for data workers.
- Ensure procedures are in place to adequately address ethical issues that may arise. This includes putting in place a process for informed consent and possibly approvals from ethics committee or institutional review boards (IRB) for data collection, depending on requirements.
- Develop checklists of all the required data collection, data entry, and data analysis supplies (e.g., pens, paper, forms, mobile phones, tablets, computers).

2. Data Collection

Data collection involves obtaining data from the original source and transferring them into tools from which they can be collated, analyzed, and reported. The tools can be paper or electronic; the transfer could involve filling out a survey form, a sign-in sheet, or a patient record.

To ensure data quality and permit multisite data comparisons, the data collection effort usually needs to be structured so that all data are collected the same way from one data collection site to the next. Credibility of the process is also very important in the data collection stage; the data must be collected in a way that yields accurate information about the actual situation. Selecting appropriate tools and methods, training data recorders and interviewers, having clear instructions on how to collect data and how to capture information, and checking the accuracy of the data generated are all part of a quality data collection process. Ensuring that data collection tools have data entry “checks” where possible will help to keep data that are collected clean. Checks can include supervisor spot-checks of data collection or data collection through mobile technology allows built-in checks on answer options and skips.

The following are some best practices for ensuring data quality during **Data Collection**:

Best Practices for Ensuring Data Quality during Data Collection

- Develop simple, streamlined, standardized, and appropriate tools, such as forms or definitions. Be mindful of language and cultural norms. Pre-test all tools.
- Ensure that methods of data collection are appropriate: Data collection should not take too long and all data collected should be used at the analysis stage.
- Carefully select data collectors, with age, marital status, and gender specified so as to best match cultural norms for target subjects (e.g., as for sexual communication and agricultural practices).
- Develop a data collection schedule. Make sure that this schedule is realistic and accounts for other work that must be done and time for return visits.
- Minimize the time between intervention activities and measurement to reduce recall problems.
- Ensure that all personnel are trained in methods, procedures, ethics, and interview techniques. As a best practice, training should include interview simulation to uncover any unclear areas of the instrument.
- Routinely check to ensure that instructions are being followed and that data are being collected correctly. Conduct spot checks to see if data collectors are entering the correct data correctly on the forms.
- Provide clear guidelines for managing the collected data (i.e., moving data from one point to the next), who is responsible and how data security is maintained.
- If revising the collection tool, identify a change management process, and follow up to make sure that people observe it.
- Ensure that data collection processes and sampling methods are consistent from year to year, location to location, data source to data source.
- Use the same instrument to collect data from year to year, location to location.
- If data come from different sources, use instruments that are similar enough that the reliability of the data is not compromised.

3. Data Collation

Data collation is the process of assembling data into a format for the purpose of analysis using either electronic or manual tools. Collation may involve data coding that labels raw data with abbreviated codes (e.g., assigning numbers to categorical variables like gender, where MALE=1 and FEMALE=2), so as to make it easier to manipulate the data.

Collation also refers to the process of consolidating information from various field sites or offices and building completed data sets. All these processes introduce an opportunity for mistakes to be made and errors to be introduced into the data. Thus, careful management of the collation process is critical to maintaining data quality. Best practice is data collation is to maintain a record or audit trail of how data are aggregated or how variables are created or re-coded. In database programs, this process is automated and in statistical packages it can be recorded by using syntax or saving procedures so that others may examine how the final variables were constructed and coded.

Streamlining Data Collation Processes

The more times data is “handled,” the more chances there are for error—every time data are reentered into a new file or spreadsheet or hand tallied, the greater the chance for error. Keep collation processes streamlined and where possible automate them to help ensure data quality.

To ensure quality during **Data Collation**, some best practices can be followed:

Best Practices for Maintaining Data Quality during *Data Collation*

- Develop checklists and signoffs for key steps as well as appropriate tools and procedures for data collation.
- Ensure that the methodology for collating the data is clear and that detailed instructions are made available to data processors. Ensure that the data collation process is well documented at each step (i.e., maintaining an audit trail).
- If using an electronic system, build formulae into the spreadsheets and clarify which source data will be aggregated and manipulated.
- Check, recheck and lock formulae in electronic systems. Use password protection with files if possible.
- Provide a schedule for data collation; include time for data clarification and verification.
- Conduct reviews during the data entry process. Actually watch how data are keyed in and assess whether checks or double data entry techniques are being used. If errors are found, communicate them to the data entry personnel.
- Routinely and randomly check for transcription errors. If necessary, double-enter data.
- Ensure that data entry and collation errors are identified during data verification.
- Identify responsible members of the team who will be entrusted with keeping and storing the data and with ensuring that data integrity and confidentiality are maintained. Assign passwords.

4. Data Analysis

Data analysis is the process of examining, reviewing, and evaluating data sets to assess a hypothesis. Analysis enables data users to test underlying hypotheses or assumptions and to predict relationships in order to understand and evaluate their programs.

The data analysis period is also an important time for a data quality review. Although the primary purpose of data analysis is to advance understanding and gain insights into the program, the analysis process also provides—from a data quality viewpoint—valuable feedback on the adequacy and completeness of our data and its relevance, validity, and precision. The data analysis process is a time when a monitoring and evaluation officer asks data quality questions such as:

- *Precision Questions:* What is the level of confidence we have in these results?
- *Validity and Relevance Questions:* What questions are not answered? (This question helps pinpoint gaps and weaknesses in the data collection process.) And did the data provide evidence to prove or disprove the matter under discussion? Are the changes that have been measured attributable at least to some extent to the organization's activities?

High quality **Data Analysis** incorporates the following procedures:

Best Practices for Ensuring Quality in Data Analysis

- Explain the analysis plan in clear simple language that all can follow. Include formulas and directions for the development of composite indicators.
- Undertake data analysis at site level, to ensure involvement of field staff in the process.
- Ensure that data is disaggregated by key variables that are essential for your program, such as gender, age, or geographical locations.
- Document and disclose all conditions and assumptions affecting interpretation of data.
- Clarify your routine analysis procedures. Will you normally compare data to targets, have standard ways you disaggregate, and/or compare geographic regions or sites? Which variables will normally be associated?
- Calculate margins of error and ensure that the margin of error is less than the expected change being measured.

5. Data Reporting

Reports are a compilation of descriptive information, presenting data analysis as useful knowledge. Reporting provides an opportunity for project implementers to inform internal and external audiences of progress, challenges, and lessons learned during implementation.

Key to the reporting process is to ensure that data presented are relevant to the different target audiences and that the report has integrity—in other words, that it accurately presents the data set and results. From a data quality perspective, accurate presentation of the findings—without an overzealous spin overstating the results or purposive exclusion of information to mislead an audience—help maintain the organization’s integrity.

Feedback mechanisms are essential to quality monitoring and evaluation processes. They provide a way to report back to the people who took the time to provide information (i.e., respondents, communities, beneficiaries, and key stakeholders), those who took the time to collect information properly (i.e., project staff and or volunteers), and those who agreed to budget for monitoring and evaluation activities (i.e., the project management staff and donors).

During **Data Reporting**, some best practices can be followed:

Best Practices for Ensuring Quality in Data Reporting

- Develop clear documentation and dissemination policies and guidelines on what information the results will include, when they will be provided, and what they will be for different audiences.
- Synthesize results for the various audiences, and present the results in a manner and format that helps each audience best understand it.
- When possible, represent data spatially or graphically so that it’s more intuitive to understand.
- Maintain integrity in reporting—do not leave out key information.
- Have multiple reviewers within the organization look at the presentation before disseminating the results, with a view to their accuracy and relevance.
- Know your data’s strengths and weaknesses and document both appropriately. Include a data limitations section if appropriate.
- Protect data confidentiality in both reports and communication tools.
- Review data with and provide feedback to those who have a stake in the results.

6. Data Usage

The ultimate goal of any monitoring and evaluation effort, *data usage* refers to the process of making timely, *data-driven decisions*. For this to happen efficiently, relevant high-quality data must be made accessible in a timely manner to key decision makers. This means that decision makers (including project teams) must know of the existence of the data set, can locate it, and can easily import it into their working environment. Usually, the data are most useful to decision makers after analysis and synthesis into an understandable, relevant report.

Although we are calling data usage the final point in the data management process, data usage takes place continuously throughout a project's life. The steps in data usage are:

- *Demand*: Program staff want data.
- *Collection*: Relevant data are gathered.
- *Availability*: Relevant data are collated, and program staff can access them.
- *Utilization*: Program staff makes decisions based on the data.

Use of available data engenders the demand for more data. Data use is enhanced if information is regarded as highly relevant to decision making among target audiences.

Dissemination policies and delivery systems, which can help determine many aspects of accessibility, are critical to overall data management.

Programs need to periodically undertake some form of explicit review to determine whether program data are actually being used for decision making and to assess whether program data are satisfying user needs—not only in terms of the range of information produced but also in terms of accuracy and timeliness.

Data use assessments typically begin with talking to those involved in project monitoring and evaluation and project management. Key questions include:

- What data are collected? What data are available?
- What do you think is the main motivating factor for most project decisions?
- What was the last project management decision to be made? What information did you use to make that decision? Was there sufficient information available?
- Have you ever had a time when data quality or availability, or the lack thereof, affected your ability to do your job?

After the project assesses its current use of data as well as the data's strengths and weaknesses, stakeholders can form a data use plan with action items to address weaknesses. Current good practices and good practices adopted to address weaknesses can form the basis of a **data use policy**.

Data Use Plan

An important part of data quality management, the *data use plan* addresses the final stage of data management, data usage. Data use plans can encourage managers to use information in decision making, provide structures or templates for documenting the use of data in decision

making, and ensure that data important to decision making are available in a timely fashion (MEASURE Evaluation 2011).

Data use plans can be flexible depending on project needs. Projects already being implemented that are just creating data use plans may first do the aforementioned data use assessment and create a plan that codifies best practices revealed by the data use assessment and action items based on data use gaps.

Projects that are starting—that is, just creating their initial monitoring and evaluation, DQM, and data use plans—should set out best practices for data use in their data use plan. A schedule for data use assessments should be included.

In some cases, a template for data use is helpful. For more templates and tools on how to better use your program data for decision making, see MEASURE Evaluation’s *Tools for Data Demand and Use in the Health Sector: Framework for linking Data with action in Annex 2*⁵ (MEASURE Evaluation, 2011). Both templates include space for a dissemination plan. As important as it is to make programmatic decisions, it is equally vital to communicate clearly about those decisions and their basis to stakeholders.

Some best practices for ensuring high quality **Data Usage** are as follows:

Best Practices for Ensuring Quality Data Usage
<ul style="list-style-type: none">• Consistently promote the use of data: Provide guidance on how data can facilitate decision making. Advocate for its use by relevant stakeholders.• Provide information that lets audiences understand what the data means and how it related to the desired changes and objectives. If applicable, write a programmatic recommendations section, along with the Programmatic Staff, based on the findings.• Routinely make data available at decision points (e.g., staff or quarterly review meetings).• Hold periodic data use assessments: Periodically undertake some form of explicit review to assess whether program data are satisfying user needs—not only as to the range of information but also in terms of accuracy and timeliness.• Utilize data use templates when major data or reports become available (i.e. after a midline evaluation) or when major decisions need to be made.

B. Data Flow Mapping

Data flow mapping is a step-by-step process where the program team constructs a graphic that documents data management processes: identification of data sources, data collection processes and tools, data collation systems and tools, data analysis, data reporting, and data use. Data flow mapping is useful for understanding how data quality, data management, and data assessment work together and is a key step in creating a DQM plan.

The map enables programs to document processes and tools that comprise data management. Visualizing a data management system, including stakeholder involvement, enables programs to better understand their systems and identify areas that need strengthening in order to avoid potential data quality problems and root out those that arise.

⁵ <http://www.cpc.unc.edu/measure/publications/ms-11-46-b>

Steps in Data Flow Mapping

- Start by outlining your current program framework for service delivery. Include a review of key program implementation processes in order to lay out the elements of data management that are key in your organization.
- Identify all stages at which important program data are sourced; note when documents and data are received.
- List all essential data collection, collation, and analysis processes, with the tools used and people responsible for managing each stage.
- List all reporting processes that your program uses to disseminate data to stakeholders.
- Identify real and potential data quality problems at various stages of data collection within your organization. Ask:

Do we have the tools for data collection, and are they appropriate?

Are data collected consistently?

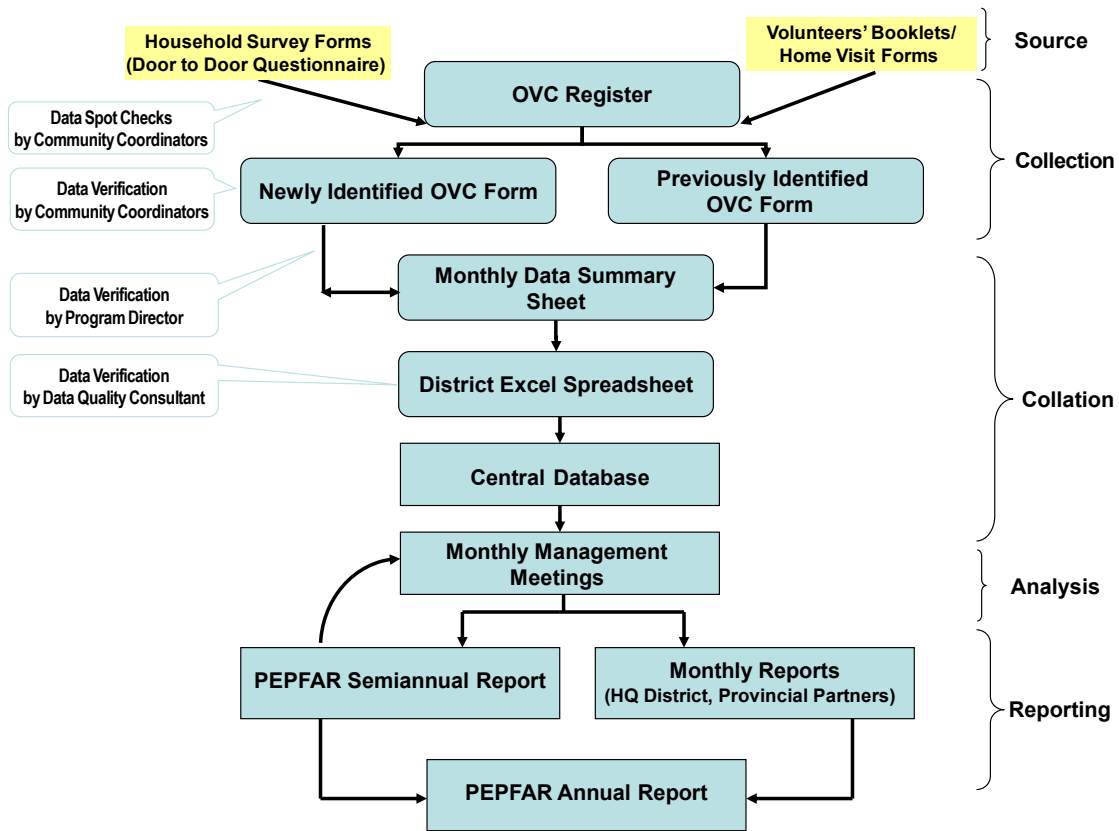
Are the responsible people properly trained on the tools for data collection?

How are the data analyzed, reported, and used for program management and or for case management and decision making?

Is data integrity maintained? Consider here data security and the integrity of respondents and data collectors.

- Brainstorm and identify solutions to address data quality problems at each stage of data collection.

Figure 1—Data flow mapping example.



EXERCISES

#3: Challenges and Solutions to Data Quality Management

Using the matrix provided on the following page, identify common problems related to the eight data quality criteria that may affect each of the six data management stages. For each criterion, determine challenges and solutions to data quality management. In this exercise, we will examine these concepts together to make sure to comprehensively address data quality issues. These solutions can be articulated in your project's PMP.

- Divide into six groups. Assign each group to work on one of the data management stages (source, collection, collation, analysis, reporting, and usage) If you do not have enough people to split into six groups, do all the stages in one group, or give each group more than one.
- Working in these groups, discuss the common problems likely to affect each of the eight DQA criteria across each of the six data management processes. Capture at least three examples for each stage of data management on the tables on the next page and note which data quality criteria it addresses (Validity, Reliability, Integrity, Precision, Timeliness, Completeness, Confidentiality, and Ethics).
- Also on these pages, brainstorm and capture *solutions* to each of the problems identified to minimize data quality problems for each assessment criteria across the six data management processes. These solutions can form the basis of SOPs, should you project decide to incorporate these throughout your PMP.

DATA MANAGEMENT PROCESS	COMMON PROBLEMS THAT MAY AFFECT DATA QUALITY	SOLUTIONS TO MINIMIZE DATA QUALITY PROBLEMS
Validity, Reliability, Integrity, Precision, Timeliness, Completeness, Confidentiality, and Ethics		

Source		
Collection		
Collation		
Analysis		
Reporting		
Data Usage		

#4: Data Flow Mapping Exercise

You will do this exercise on your own.

- Choose one program from your organization.
- Draw a graphic like the Figure 1 Data Flow Map in this chapter outlining what occurs at each step of the data management process for that entire program: source, collection, collation, analysis, reporting, and usage.
- Make note of any data quality problems you see at each stage.
- Write possible solutions to those data quality problems.

Chapter 4: Data Quality Assessments and Audits

In this chapter, readers will learn:

- The rationale and approaches for data quality assessment.
- Data quality assessment and audit procedures.

In building a data quality management plan, readers will:

- Outline the process for undertaking RDQA for their program (Exercise 5).
- Identify key stakeholders and their role in the RDQA process (Exercise 5).

The Data Quality Assessment Process

The DQA is an essential procedure that allows an organization or donor to determine data quality at any given time and the opportunity to develop and implement strategies to address and prioritize gaps. The process consists of asking pointed questions on data quality and data management processes and researching the answers. By asking these questions, the organization can determine a data set's potential for error and therefore understand how confident the staff can be with the results and in using the data to evaluate the program and make management decisions.

Routine data quality assessments (RDQAs) are similar to data quality assessments, except while DQAs are externally conducted, usually by the donor agency, RDQAs are conducted by the project. There is more room for flexibility, more leeway to adapt to the organization's needs rather than the donor's requirements. For guidance on how frequently to conduct RDQAs and with what quantity of partner organizations, please view the latest version of the Pact Results and Measurement Standards on the R&M Mosaic page.⁶

Like DQAs, RDQAs have three primary components:

1. *Data Management Review*: Are the data management systems and structures in place adequate to ensure data quality?
2. *Data Verification/Indicator Assessments*: Are the data being collected accurate?
3. *Developing a Data Quality Action Plan*: If there are problems with data verification or data management, how should the organization proceed? What areas should be prioritized for improvement? Who should be responsible for following through on these actions?

1. Data Management Review

- Data management review enables the reviewer to identify any gaps that may affect data quality.

Data management review can also be a good time for data flow mapping (See Chapter3).

⁶ *Pact Quality Standards for Results and Measurement*. <https://pactworld.jiveon.com/groups/results-and-measurement>

2. Data Verification/Indicator Assessments

Indicator assessments determine whether there are issues with data quality as assessed using the VRIPT-CCE criteria discussed in Chapter 1 (validity, reliability, integrity, timeliness, and precision, plus completeness and confidentiality and ethics). To complete the assessment, you ask a series of questions and research the answers (*see instructions on the RDQA Tool in Annex I*). By asking data quality questions, the organization can determine weaknesses in the data management system and areas that may be in need of strengthening.

After careful comparison and mapping of various RDQA tools, Pact has chosen to adopt a slightly modified version of Measure Evaluation’s RDQA Tool⁷. It is a straight-forward, easy to use tool that allows for assessments at the national level (Pact Country Office), regional level (if appropriate), district level, and partner level (or “Service Delivery Site,” as it is known in the tool). In the Excel RDQA Tool, each question is assessing one or more of the VRIPT-CCE data quality criteria.⁸ Between 30 and 35 questions are asked in two parts:

Part 1) Verification of reported data for key indicators (up to four indicators)

Part 2) Assessment of the data management and reporting systems.

Part 2 is divided into five functional areas:

1. M&E Structure, Function & Capabilities
2. Indicator Definitions and Reporting Guidelines
3. Data Collection and reporting Forms/Tools
4. Data Management Processes
5. Links with the National Reporting System

After answering a series of questions, the tool provides a dashboard of results that allow the program team to prioritize action areas. The tool is in an Excel file, but could also be completed on paper. Instructions on how to use the tool can be found in **Annex 1**.

3. Developing a Data Quality Action Plan

Following the data quality assessment, the program team should develop a strategy for strengthening data quality, with clear objectives. The dashboard at the bottom of each service delivery site tab contains spider and bar graphs showing what areas are weaker than others. Based on your analysis of scores, an action plan for each partner or site should be developed. An action plan is the set of data quality strengthening objectives (DQSOs), actions, due dates, and who is responsible on a program team in addressing the identified data quality issues. Ideally, the program team selects activities likely to result in the biggest improvements in data quality. Representative DQSOs include:

- Improving implementation of indicator protocols at all program sites.
- Improving field data management systems via on-site support and mentoring.
- Adequately training data collectors.

⁷ Original tool: <http://www.cpc.unc.edu/measure/tools/monitoring-evaluation-systems/data-quality-assurance-tools>

For the latest Pact-modified RDQA Tool, please visit: <https://pactworld.jiveon.com/groups/results-and-measurement>

⁸ Annex 1 has the complete mapping of each question to the VRIPT-CCE criteria, as does a tab in the Excel file called ‘List of Questions.’

- Conducting routine data verification and validation processes for high-risk partners /sites (partners /sites with low average scores across each of the five functional areas).

Procedures for Conducting an RDQA of Grantee or Partner Data

In situations where an organization provides a grant to a partner organization (a common element of Pact projects) the partners may collect data that are then used by the grantor for program analysis and evaluation. Any data used for decision making must be accurate. Data collected under a grant program are considered official and auditable, and the partner is held accountable, just as for financial data. To assess partner data, the procedures are as follows:

- An office visit is conducted to meet with the partner. During this meeting, the RDQA process is explained and general information obtained on the partner’s data management procedures. Part 2, the Systems Assessment portion of the “Service Delivery Site” tab in the Excel tool can be used for this purpose. **For Pact, this visit may happen during the pre-award phase or when an Organizational Capacity Assessment (OCA) or Organizational Performance Index (OPI) measurement is being made.**
- A second visit is undertaken (or a series of visits) to begin verification and validation of pre-selected indicators. If the project has a large number of partners, it may be necessary to draw a sample of partners with which to conduct the RDQA.
- On **critical target indicators** (e.g., number of people served), **an RDQA is conducted** in a standard RDQA format. In the Excel tool, this means completing both section 1 and 2 in the sheet. Findings are then documented, with notes on the various issues. Once the RDQA is complete, recommendations should be developed, based on lower performing areas.
- Meet again with the partner organization and provide feedback on the findings and recommendations. Together, plan the way forward and develop a system strengthening plan.
- Follow up on recommended actions to ensure they are implemented. Double-check field system functionality via site visits and discussions with target beneficiaries.

The Data Auditing Process

Data audits are similar in process to data quality assessments, but vary in one critically important element. The major focus of an **assessment** is to **determine risk and develop a plan of action to reduce risk**. The major focus of an **audit** is to **determine whether or not the data we are providing** (or that are being provided to us) **are accurate**. Often a data quality audit is undertaken by a funding agency.

For example, after we reported that the number of people reached with a service was 1,354, an auditor might be asked to confirm that number using our records. Is the number really 1,354, or was it 2,000, or 200, or is the data of

The ability of an organization to maintain a verifiable audit trail is the fundamental test of the maturity of its data management system.

such poor quality that the auditor cannot confirm the number with any precision?

Thus, the purpose of a data quality audit is to ascertain, on the basis of *objective evidence*, whether or not data reported by the auditee meets minimum acceptable standards of data quality. Of note, RDQAs can be good preparation for an audit.

Data are audited based on a set of criteria determined by the people or team that commissions the audit. However, audits commonly include assessment of data quality (i.e., validity, reliability, integrity, precision, timeliness, as well as data completeness, confidentiality and ethics) in light of the data management systems by which such data are generated.

The approach to data quality auditing may also vary depending on the requirements set by the audit commissioner. One common approach is the three-phase process often used by the United States Government (USG). This typically includes some capacity building for the auditee in addition to establishing the quality of existing data. This audit takes approximately one year to complete and is described in detail below. Each phase of the program has a particular function, but the three phases are very much linked to one another.

Phase 1

This phase of the audit cycle is an evaluation of the partner/grantee's entire data management system and the associated processes and procedures—that is, an evaluation of the partner/grantee's ability to produce the required data at the required level of quality. Partners/grantees who report large volumes of data and/or who receive substantial funding are usually targeted during this phase. High-risk data handling practices may be identified during this phase, giving the partner/grantee a chance to improve practices before Phase 2 and the next reporting cycle.

Phase 2

This is an actual verification and validation of a specific predetermined data set, with attention to validity, reliability, integrity, precision, timeliness, completeness, and confidentiality and ethics. The *total risk scores* assigned during the audit (that is, the cumulative probability of error considering all relevant factors within a data set) are used to determine the necessity for issuing *compliance notes* and the extent of assistance the partner/grantee may need. A compliance note includes evidence that the data management was of insufficient rigor as well as recommended actions designed to prevent recurrence of the poor data practices. Compliance notes are issued in either a minor or major category,

Leaving an Audit Trail

An audit trail is important to data quality. The collection of documents related to the data, it includes the original data source and notes on all data aggregation, tallying, editing, and summarizing that occurs. It also includes documents that illustrate that results are reported in the same manner over time to ensure reliability and that measurement and/or instrumentation error are minimized. A data audit trail helps ensure that both the DQA team and the donor understand exactly how the results are derived. As part of its own SOP, Pact maintains audit trails and mentors partners and grantees in doing the same.

based on the DQA score. A score of between 12 and 16 usually results in a *major compliance note*, while lower scores engender a *minor compliance note*. Auditors usually issue compliance notes only for scores of nine and above.

Phase 3

This phase of the audit cycle—often called the *closeout* of a compliance note—is a review of the compliance notes issued to partners/grantees during Phase 2, in order to verify that the partner/grantee implemented the actions

recommended to address the identified data quality issues. Other data are also verified to ensure that poor data practices have not affected other data sets, especially when major compliance notes were issued. When there are outstanding issues, they are brought to the attention of the audit commissioner for further action.

Following each phase of the audit, the auditor prepares a detailed report documenting the audit’s process, outcome, and recommendations. These reports are submitted to the commissioning agency team, as well as to the auditee for review and action.

One of the most important things you can do to prepare for data audits and data quality assessments is to be familiar with your donor’s requirements and expectations. Donors such as USAID and the Global Fund have strict requirements for both assessments and audits. On the other hand, many European donors provide few guidelines, if any. All M&E staff must understand donor-stipulated regulations. Links to donor requirements follow:

USAID

USAID. 2010. Performance monitoring and evaluation tips: Conducting data quality assessments. *Tips*, no. 18, 1st edition.

http://pdf.usaid.gov/pdf_docs/pnadw118.pdf

CIDA

Foreign Affairs, Trade and Development Canada. 2012. Data integrity control framework.

<http://www.acdi-cida.gc.ca/acdi-cida/acdi-cida.nsf/eng/NAT-111672020-GGS>.

Global Fund

The Global Fund to Fight AIDS, Tuberculosis, and Malaria. 2013. Data quality tools and mechanisms. <http://www.theglobalfund.org/en/me/documents/dataquality/>.

Pay attention to compliance notes, and fix these problems before the auditor returns.

Failure to do so can result in losing your cooperative agreement or grant!

EXERCISE

#5: Identifying Key Data Quality Assessment Steps for Your Program

Spend 30 minutes working individually or in small groups to list a potential step-by-step RDQA process based on your program management and implementation processes. Consider activities that would need to be implemented in the field versus those at the head office, as well as the potential problems that are likely to exist in the current data management system. Also consider the roles of different stakeholders in implementing this process

The RDQA process should include the following key steps:

KEY INDICATORS	PARTIES RESPONSIBLE FOR COLLECTING DATA	FREQUENCY OF RDQA
1.		
2.		
3.		
4.		

Key stakeholders to be included and their role in the RDQA process are:

KEY STAKEHOLDERS	ROLE IN THE ROUTINE DATA QUALITY ASSESSMENT PROCESS
1.	
2.	
3.	
4.	

Appendix 1: How to Use the Routine Data Quality Assessment Tool

Including Instructions on Developing Action Plans

After careful comparison and mapping of various RDQA tools, Pact has chosen to adopt Measure Evaluation's RDQA Tool. This tool was created with input from a consortium of organizations including the Global Fund to Fight Aids, Tuberculosis and Malaria, Office of the Global AIDS Coordinator, PEPFAR, USAID, WHO, and UNAIDS.⁹ It is a simple, straight-forward tool that can be used across many types of programs at different levels of a program. The tool is intact except for a few questions Pact added to cover validity, precision and ethics and Pact also added the ability to assess up to four indicators per partner.¹⁰ Additionally, some of the terminology was changed to ensure that the tool can be widely understood by both health and non-health programs.

Of note, most of the instructions in this Appendix (with only minor modifications) were taken directly from Measure Evaluation's "*Routine Data Quality Assessment tool: Guidelines for Implementation.*" Full credit must be given to Measure Evaluation for the text.¹¹

Overview of the Tool

The Routine Data Quality Assessment (RDQA) Tool allows programs and projects to rapidly assess the quality of their data and strengthen their data management and reporting systems.

The objectives of the RDQA Tool are to:

- VERIFY rapidly 1) the quality of reported data for key indicators at selected sites; and 2) the ability of data-management systems to collect, manage and report quality data.
- DEVELOP an action plan to implement corrective measures for strengthening the data management and reporting system and improving data quality.
- MONITOR over time capacity improvements and performance of the data management and reporting system to produce quality data (notably through repeat use of the RDQA).

This tool may be used to develop an initial data quality plan and also when conducting periodic data quality assessments and determining action areas in need of improvement.

The RDQA exists in Microsoft Excel format. The checklists can be printed and completed by hand or, alternately, responses can be entered directly into the spreadsheets on a computer.

⁹ Original tool: <http://www.cpc.unc.edu/measure/tools/monitoring-evaluation-systems/data-quality-assurance-tools>

¹⁰ For the latest Pact-modified RDQA Tool, please visit: <https://pactworld.jiveon.com/groups/results-and-measurement>

¹¹ http://www.cpc.unc.edu/measure/tools/monitoring-evaluation-systems/data-quality-assurance-tools/RDQA%20Guidelines_2010.doc

When completed electronically, a number of dashboards produce graphics of summary statistics for each site or level of the reporting system, as well as a global dashboard that aggregates the results from all levels and sites included in the assessment.

The RDQA Tool allows Pact to assess data quality and management at various levels of the data reporting system: the M&E Unit (or Pact’s central country program); a regional or intermediate aggregation site; and the Service Delivery Point, **which for Pact’s purposes will usually be the partner organization.**

For many Pact projects, only Service Delivery Sites (Partners) and the National M&E Unit (Pact’s Central Country Office) may be relevant.

Elements of the Tool

The tool consists of two parts:

1. Part 1 is the data verification exercise for a selected indicator (s).
2. Part 2 is the M&E systems assessment.

Part 1 of Tool - Verification of Reported Data for Key Indicators:

The purpose of Part 1 of the RDQA is to assess, on a limited scale, if service delivery and intermediate aggregation sites are collecting and reporting data to measure the audited indicator(s) accurately and on time — and to cross-check the reported results with other data sources. To do this, the RDQA will determine if Service Delivery Sites / Partners have accurately recorded the activity related to the selected indicator(s) on source documents. **Of note, you can select a sample of partner sites or conduct with all partners, if feasible. Please reference the latest Pact Results and Measurement Standards for guidance on what proportion of partners to include and how frequently to conduct the RDQAs.**

The RDQA will then trace that data to see if it has been correctly aggregated and/or otherwise manipulated as it is submitted from the initial Service Delivery Sites/Partners through intermediary levels to the program/project M&E Unit. **If a critical project indicator is selected for the RDQA and we conduct an RDQA with a selection of partners (or all partners), the Service Site Dashboard will aggregate all scores for that indicator across partners. You can review up to four indicators per partner/level.**

Data Verifications at Service Delivery Points (usually Pact partners):

At the Service Delivery Points, the data verification part of the RDQA Tool (Part 1) has three sub-components:

1. Reviewing Source Documentation: Review availability and completeness of all indicator source documents for the selected reporting period.
2. Verifying Reported Results: Recount the reported numbers from available source documents, compare the verified counts to the site reported number; and identify reasons for differences.

3. Cross-checking Reported Results with other Data Sources: Perform cross-checks of the verified report totals with other data-sources (e.g. inventory records, laboratory reports, registers, etc.).

Data Verifications at Intermediate Aggregation Sites and at the National M&E Unit (Pact Country Office):

At the Intermediate Aggregation Sites and the National M&E Unit (for Pact purposes, this would be the Pact Country Office), the data verification part of the RDQA Tool (Part 1) has two sub-components:

1. Reviewing Site Reports: Review availability, timeliness, and completeness of expected reports from Service Delivery Sites for the selected reporting period.
2. Verifying Reported Results: Re-aggregate the numbers from the reports submitted by the Service Delivery Points, compare the verified counts to the numbers submitted to the next level (e.g; M&E Unit), and identify reasons for any differences.

Part 2 of Tool- Assessment of Data Management and Reporting Systems:

The purpose of Part 2 of the RDQA is to identify potential threats to data quality posed by the design and implementation of the data management and reporting system at three levels: (1) the program/project M&E Unit, (2) the Service Delivery Points, and (3) any Intermediary Aggregation Site (e.g. district or regional offices) at which reports from Service Delivery Points are aggregated prior to being sent to the program M&E Unit (or other relevant level). The questions for the systems assessment are grouped in the following five functional areas:

1. M&E Structures, Functions and Capabilities
2. Indicator Definitions and Reporting Guidelines
3. Data Collection and Reporting Forms and Tools
4. Data Management Processes
5. Links with National Reporting System—**of note, this section may not be**

relevant for many types of Pact programs and may be skipped. The exception is health programs where there may be national reporting requirements for which you are required to report health data.

Interpretation of the Dashboard Spider-Graphs

The scores generated for each functional area on the spider graph are an average of the responses which are coded 3 for “Yes, completely”, 2 for “Partly” and 1 for “No, not at all.” Responses coded “N/A” or “Not Applicable” are not factored into the score. The numerical value of the score is not important; the scores are intended to be compared across functional areas as a means to prioritizing system strengthening activities. That is, the scores are relative to each other and are most meaningful when comparing the performance of one functional area to another.

For example, if the system scores an average of 2.5 for ‘M&E Structure, Functions and Capabilities’ and 1.5 for ‘Data-collection and Reporting Forms/Tools’ one would reasonably conclude that resources would be more efficiently spent strengthening ‘Data-collection and Reporting Forms/Tools’ rather than ‘M&E Structure, Functions and Capabilities.’ The scores should therefore not be used exclusively to evaluate the data management and reporting

system. Rather, they should be interpreted within the context of the interviews, documentation reviews, data verifications and observations made during the RDQA exercise.

Additional Guidance on the Tool

- Parts 1 (data verifications) and 2 (systems assessment) of the RDQA Tool can be implemented at any or all levels of the data management and reporting system: M&E Unit; Intermediate Aggregation Sites; and/or Service Delivery Points/ Partner Level. While it is recommended that both parts of the RDQA be used to fully assess data quality, especially the first time it is being implemented; depending on the assessment objectives, one or both of these protocols can be applied and adapted to local contexts. **It is however recommended that the data verification part of the tool (with the select indicator) be conducted more frequently in order to monitor and guarantee the quality of reported data. The system assessment protocol could be applied less often (e.g., annually).**
- It is recommended to apply the Part 2 System Assessment in a **participatory manner** with all relevant M&E and program staff present and discussing the answers thoroughly. As questions are answered, detailed notes should be taken to ensure a comprehensive understanding of the responses. This is also necessary so that follow-up visits can ensure the correct improvements have been made.
- If any of the questions are not relevant, please select N/A. After completing the questions in the Excel tool, a dashboard will present the findings. From the findings, you should develop an Action Plan.
- The Pact version of the tool also contains a final column that provides additional guidance to the evaluator conducting the RDQA.

Description of Each Sheet in the Tool

The RDQA Tool consists of a series of spreadsheets in a Microsoft Excel file:

Sheet 1- START: to select the number of service delivery points / partners and intermediate aggregation sites (e.g. district or regional offices, if applicable) to be included in the RDQA. Selecting these numbers automatically provides you with the number of tabs you need for your assessment.

For Pact’s purposes, the “Service Delivery Site” will often refer to a Pact partner. The “Regional Aggregation Sites” and “District Aggregation Sites” are only applicable if the data is passing through these intermediary steps before reaching the “National Level M&E Unit”, which refers to the Pact Country Office. In most types of programming, we would only use the Service Delivery Level and National M&E Unit level tabs for the RDQA.

Sheet 2- Instructions: to inform users how to use the Excel spreadsheet.

Sheet 3- Information: to record the country, program/project, indicator and reporting period

assessed, documentation reviewed and composition of the assessment team.

Sheet 4- Service Delivery Point:

- to record the results of the 1) data verifications, and 2) system assessment at the service site;
- to display a dashboard of results for the service site (more detail provided below);
- to capture recommendations for the service site.

As previously mentioned, if conducting an assessment of the same indicators across different partners, use one Excel file but each partner/site will have its own sheet within the file.

Sheet 5- Intermediate Aggregation Site:

- to record the results of the 1) data verifications, and 2) system assessment at the intermediate site;
- to display a dashboard of results for the intermediate site;
- to capture recommendations for the intermediate site.

Sheet 6- National M&E Unit / Pact Country Office:

- to record the results of the 1) data verifications, and 2) system assessment at the M&E Unit;
- to display a dashboard of results for the M&E Unit;
- to capture recommendations for the M&E Unit.

Sheet 7- Global Dashboard:

- To display in a graphic format the aggregated results from all sites and levels visited in the RDQA.

Sheet 8- RDQA Final Action Plan:

- To consolidate recommendations from each level into an overall action plan based on the RDQA,

Sheet 9- List of Survey Questions:

- A reference page to map the functional areas assessed in the systems assessment (dimensions of data quality) part of the RDQA with components of data quality. The “List of Survey Questions” tab lists the questions asked for the systems assessment, the levels of the reporting system to which the questions pertain, and the components of data quality addressed by each question.

Steps for Conducting the RDQA

The steps below can be found on the Introduction tab of the Excel tool, but are also copied below for your reference.

Steps for Conducting the RDQA

1. Determine Purpose:

The RDQA checklist can be used for:

- Initial assessment of M&E systems established by new implementing partners (or in decentralized systems) to collect, manage and report data.
- Routine supervision of data management and reporting systems and data quality at various levels. For example, routine supervision visits may include checking on a certain time period worth of data (e.g. one day, one week or one month) at the service site level, whereas periodic assessments (e.g. quarterly, biannually or annually) could be carried out at all levels to assess the functioning of the entire Program/project's M&E system.
- Periodic assessment by donors of the quality of data being provided to them (this use of the DQA could be more frequent and more streamlined than official data quality audits that use the DQA for Auditing) but less frequent than routine monitoring of data.
- Preparation for a formal data quality audit.
- The RDQA is flexible for all of these uses. Countries and programs are encouraged to adapt the checklist to fit local program contexts.

2. Level/Site Selection:

Select levels and sites to be included (depending on the purpose and resources available). Once the purpose has been determined, the second step in the RDQA is to decide what levels of the data-collection and reporting system will be included in the assessment - service sites, intermediate aggregation levels, and/or central M&E unit. The levels should be determined once the appropriate reporting levels have been identified and "mapped" (e.g., there are 100 sites providing the services in 10 districts. Reports from sites are sent to districts, which then send aggregated reports to the M&E Unit). In some cases, the data flow will include more than one intermediate level (e.g. regions, provinces or states or multiple levels of program organizations). **However, intermediate sites may not always be relevant for all types of Pact programs and could be omitted.**

3. Identify indicators, data sources and reporting period:

The RDQA is designed to assess the quality of data and underlying systems related to indicators that are reported to programs or donors. It is necessary to select one or more indicators – or at least program areas – to serve as the subject of the RDQA. **You may select the indicator because it is one of the most critical indicators of the project, most partners are reporting on it, etc.**

4. Conduct site visits:

During the site visits, the relevant sections of the appropriate checklists in the Excel file are filled out (e.g. the service site checklist at service sites, etc). These checklists are completed following interviews of relevant staff and reviews of site documentation. Using the drop down lists on the START page of the workbook, select the appropriate number of Intermediate Aggregation Levels (IAL) and Service Delivery Points (SDP) to be reviewed. The appropriate number of worksheets will automatically appear in the RDQA workbook (up to 12 SDP and 4 IALs).

5. Review outputs and findings:

The RDQA outputs need to be reviewed for each site visited. Site-specific summary findings in the form of recommendations are noted at each site visited.

The RDQA checklists exist in MS Excel format and responses can be entered directly into the spreadsheets on the computer. Alternatively, the checklists can be printed and completed by hand. When completed electronically, a dashboard produces graphics of summary statistics for each site and level of the reporting system.

The dashboard displays two (2) graphs for each site visited:

- A spider-graph displays qualitative data generated from the assessment of the data-collection and reporting system and can be used to prioritize areas for improvement.
- A bar-chart shows the quantitative data generated from the data verifications; these can be used to plan for data quality improvement.

In addition, a 'Global Dashboard' shows statistics aggregated across and within levels to highlight overall strengths and weaknesses in the reporting system. The Global Dashboard shows a spider graph for qualitative assessments and a bar chart for quantitative assessments as above. In addition, strengths and weakness of the reporting system are displayed as dimensions of data quality in a 100% stacked bar chart. For this analysis questions are grouped by the applicable dimension of data quality (e.g. accuracy or reliability) and the number of responses by type of response (e.g. 'Yes - completely', 'Partly' etc.) are plotted as a percentage of all responses. A table of survey questions and their associated dimensions of data quality can be found on the 'Dimensions of data quality' tab in this workbook.

6. Develop a system's strengthening plan, including follow-up actions:

The final output of the RDQA is an action plan for improving data quality which describes the identified strengthening measures, the staff responsible, the timeline for completion, resources required and follow-up. Using the graphics and the detailed comments for each question, weak performing functional areas of the reporting system can be identified. Program staff can then outline strengthening measures (e.g. training, data reviews), assign responsibilities and timelines and identify resources using the Action Plan tab in the workbook.

Mapping of Questions in Tool to Corresponding VRIPT-CCE Data Quality Criteria

Each question in the tool relates to a data quality dimension. This information is in a tab in the Excel tool, but also copied here so you may see how the question corresponds to the VRIPT-CCE dimensions we have learned about.

Systems Assessment Components Contributing to Data Quality Dimensions												
Functional Area	Level			Dimension of Data Quality								
	M&E Unit	Aggregation Levels	Service Points	Accuracy	Reliability	Timeliness	Completeness	Precision	Confidentiality	Integrity	Ethics	Validity
<i>I - M&E Structure, Functions and Capabilities</i>												
There is a documented organizational structure/chart that clearly identifies positions that have data management responsibilities at the M&E Unit. (to specify which Unit: e.g. MoH, NAP, GF, World Bank)	✓			•	•	•						
All staff positions dedicated to M&E and data management systems are filled.	✓			•	•	•						
A senior staff member (e.g., the Program Manager) is responsible for reviewing the aggregated numbers prior to the submission/release of reports from the M&E Unit.	✓			•	•		•	•				
There are designated staff responsible for reviewing the quality of data (i.e., accuracy, completeness, timeliness and confidentiality) received from sub-reporting levels (e.g., regions, districts, service points).	✓	✓		•	•	•	•	•	•			
There are designated staff responsible for reviewing aggregated numbers prior to submission to the next level (e.g., to the central M&E Unit).		✓	✓	•	•							
The responsibility for recording the delivery of services on source documents is clearly assigned to the relevant staff.			✓	•	•							
There is a training plan which includes staff involved in data-collection and reporting at all levels in the reporting process.	✓			•	•	•	•		•			
All relevant staff have received training on the data management processes and tools.	✓	✓	✓	•	•	•	•	•	•			

II- Indicator Definitions and Reporting Guidelines											
The M&E Unit has documented and shared the definition of the indicator(s) with all relevant levels of the reporting system (e.g., regions, districts, service points).	✓			•	•						
There is a description of the services that are related to each indicator measured by the Program/project.	✓			•	•						
The M&E Unit has provided written guidelines to all reporting entities (e.g., regions, districts, service points) on reporting requirements and deadlines.	✓	✓	✓	•	•	•	•				
There is a written policy that states for how long source documents and reporting forms need to be retained.	✓			•	•	•	•	•		•	•
III- Data-collection and Reporting Forms / Tools											
If multiple organizations are implementing activities under the Program/project, they all use the same reporting forms and report according to the same reporting timelines.	✓			•	•						
The M&E Unit has identified a standard source document (e.g., medical record, client intake form, register, etc.) to be used by all service delivery points to record service delivery.	✓			•	•						
The M&E Unit has identified standard reporting forms/tools to be used by all reporting levels	✓	✓	✓	•	•						
...The standard forms/tools are consistently used by all levels.	✓	✓	✓	•	•						
Clear instructions have been provided by the M&E Unit on how to complete the data collection and reporting forms/tools.	✓	✓	✓	•	•						
The data collected by the M&E system has sufficient precision to measure the indicator(s) (i.e., relevant data are collected by sex, age, etc. if the indicator specifies disaggregation by these characteristics).	✓		✓					•			
[IF APPLICABLE] The margin of error is larger than the change being reported (and is reported).	✓		✓					•			
[IF APPLICABLE] The sampling frame (i.e., list of units in the target population from which the sample is selected) must be up to date, comprehensive, and mutually exclusive for separate geographic areas. [Measurement Validity]	✓		✓								•

If data discrepancies have been uncovered in reports from sub-reporting levels, the M&E Unit (e.g., districts or regions) has documented how these inconsistencies have been resolved.	✓	✓		•	•	•	•	•		•	•	•
The M&E Unit can demonstrate that regular supervisory site visits have taken place and that data quality has been reviewed.	✓			•	•	•	•	•	•	•	•	•
V- Links with National Reporting System												
When available, the relevant national forms/tools are used for data-collection and reporting.	✓	✓	✓	•	•			•		•	•	•
When applicable, the data are reported through a single channel of the national reporting system.	✓	✓	✓	•	•			•		•	•	•
Reporting deadlines are harmonized with the relevant timelines of the National Program (e.g., cut-off dates for monthly reporting).	✓	✓		•	•			•		•	•	•
The service sites are identified using ID numbers that follow a national system.	✓	✓		•	•			•		•	•	•
The system records information about where the service is delivered (i.e. region, district, ward, etc.)	✓	✓	✓	•	•			•		•	•	•
...if yes, place names are recorded using standardized naming conventions.	✓	✓	✓	•	•			•		•	•	•

Appendix 2: Data Quality Management Plan TEMPLATE

ORGANIZATION NAME

DATA QUALITY MANAGEMENT PLAN

PROGRAM NAME

DATE UPDATED

SUMMARY SCORES FROM RDQA

How to Use This Template

In the text of this section, *italics type* denotes topics that are suggested for coverage; normal, non-italicized type denotes language that you might adopt verbatim for your data own quality management plan.

Data Quality Management Plan Template: Table of Contents

- A. Introduction & Overview
 - A1. Introduction to Purpose of DQM Plan
 - A2. Overview of key indicators & key stakeholders**
- B. Description of the Data Management System
 - B1. Data Management Process (Data Use Plan included here)
 - B2. Data Flow Map

- C. Description of Routine Data Quality Assessments
 - C1. Routine Data Quality Assessment Process & Frequency of Assessments
 - C2. Description of Tool
 - C3. Site Visit Procedures

- D. Reports and Action Plans
 - D1: Data Quality Dashboard Report
 - D2: Action Plans

A. Introduction & Overview

A1. Introduction to Purpose of DQM Plan

[Provide a brief overview of the purpose of this plan; the value it adds to the organization’s MERL systems; key aspects of the MERL system it addresses; and who should use the plan and how.]

A2. Overview of key indicators & key stakeholders

[List here key indicators of the project, for example indicators that are being reported across multiple partners or indicators of particular importance to the donor]

[Insert table of stakeholders from Exercise 6 here]

Key stakeholders to be included and their role in the RDQA process are:

KEY STAKEHOLDERS	ROLE IN THE ROUTINE DATA QUALITY ASSESSMENT PROCESS
1.	
2.	
3.	
4.	

B. Description of the Data Management System

B1. Data Flow Map

[Please insert the data flow map created in Exercise 5 here]

B2. Data Use Plan

The purpose of this data use plan is to map all of the data sources the project collects, how those data are used in the project, and who sees the results of those data. This map will allow project management and technical staff to be better aware of what data sources are available to make informed decisions and to describe project achievement. It is also a learning exercise for M&E staff to be more aware of how the data the project collects is actually used once it is in its final form, and for the project as a whole to identify gaps or room for growth in data collection. This includes reevaluating whether all data sources are actually useful for managers (those that are not should be cut or revised) and understanding how different sectors and partners collect, use and disseminate data, and how those methods may or may not be replicated.

The inputs to this data use plan were initially developed **DATE**. This document was last updated **DATE** and is a living document that will be updated regularly to reflect changing data sources and uses.

[Please select either Version 1 or Version 2 of the Data Use Plan. Version 1 is a Data Use Plan organized by data source. Version 2 is a Data Use Plan organized by research question].

Version 1: Data Use Plan by Data Source

Data Source	Format	Frequency	Audience	Dissemination Channel	Primary purpose	How specifically is it used?
Baseline/ Midterm survey	SPSS	Baseline, midterm, endline	External stakehold ers, donor	Final report is presented to USAID, posted on external website, published in hard copy. Results also used in APHA presentation and other external presentations.	Evaluation	To understand whether the project is having an impact and to adjust program activities to findings.
Objective 1: Decrease child, maternal and newborn mortality						
Pact MCH						
Monthly project report	Excel	Monthly	Managers , M&E	Included in quarterly	Managem ent	Managers look at the data against targets and

Data Source	Format	Frequency	Audience	Dissemination Channel	Primary purpose	How specifically is it used?
			<i>and Program Team</i>	<i>report. Program team gets the summary data and send feedback to township team.</i>		<i>use the information to make any necessary program changes. The data on participants give information on community participation and coverage that inform plans for future workshops, such as minimum required number of participants.</i>
Objective 2:						
<i>Pre/Post Survey</i>	<i>Magpi</i>	<i>Before and After Self-learning Materials Reading Session</i>	<i>PC, Tech Lead, Program Team, M&E and Township Managers</i>	<i>Quarterly report to donors, results are included in external presentations</i>	<i>Assessing volunteers' knowledge</i>	<i>Managers figure out the increased knowledge of the community with internal management in order to make any necessary program changes. Pretest data show less awareness part of the community to design the program workshop and training to cover these areas.</i>

Version 2: Data Use Plan Organized by Research Question

From MEASURE Evaluation’s Tools for Data Demand and Use in the Health Sector: Framework for linking Data with action¹² (Template Version 1, Page 6)

TITLE:					
OBJECTIVE:					
TIME PERIOD:					
RESEARCH QUESTION	FINDINGS	RECOMMENDATIONS FOR ACTION/ DECISION	DECISION MAKER (DM) AND OTHER STAKEHOLDERS (OS)	COMMUNICATION CHANNEL TO REACH DM AND OS	TIMELINE

¹² <http://www.cpc.unc.edu/measure/publications/ms-11-46-b>

C. Description of Routine Data Quality Assessments

- C1. Routine Data Quality Assessment Process & Frequency of Assessments
- C2. Description of Tool
- C3. Site Visit Procedures

C1. Routine Data Quality Assessment Process & Frequency of Assessments

We conduct RDQAs of key program indicators [*enter frequency here, e.g. at least once annually*] to determine the worthiness of data for reporting and programmatic decision making. The RDQA process focuses on asking questions on the data management process as well as checks the accuracy of data reported. Routine data quality management (RDQM) is implemented at different levels of the program (e.g., head office, regional office, site), thereby enabling the identification of challenges at different levels.

The RDQA process can serve several purposes including:

- An initial assessment of M&E systems established by new implementing partners or at a new site (Using Part 2 of the ‘Service Delivery Site’ tab of the RDQA Excel Tool)
- Part of a routine supervision process for individuals working at lower levels of the data management system (e.g., at site or regional level).
- Periodic assessment by donors of the quality of data being provided to them.
- Preparation for a formal data quality audit.

C2. Description of the RDQA Tool

The Excel-based RDQA tool consists of a set of questions intended to generate responses reflective of the status of the data management systems and data quality at the different levels at which data are handled. The tool addresses the various data quality and management criteria.

C3. Description of Site Visit Procedures

Planning for Routine Data Quality Assessments, Using Results to Improve Processes

In planning for the RDQA, steps are taken to ensure active participation of different key stakeholders in the organization, particularly those involved in management and decision-making processes at different levels. Ensuring active participation of decision makers enhances buy-in and the ability to use findings from the RDQA to make decisions. When this is undertaken as part of routine supervision processes, we select sites based both on observations made during routine reporting and on the potential risk for data quality problems at the different levels of the data management system.

Results from the routine data quality management present an overview of the state of data quality at different levels of the MERL system. At each site where the tool is implemented, a report is generated highlighting the status of data quality based on responses to questions and findings from the trace and verification exercise. The tool also provides a means of aggregating the data and thereby generating a “global dashboard” that presents the overall status of data management systems across the organization/Program.

Review of MERL Systems' Capacity at Partner/Site Level

In cases where our program works through community-based partners to implement activities and to report accordingly, it is essential that MERL systems' capacity is systematically assessed. Data collected under a grant program is considered official and auditable information for which the organization is held accountable (just like financial data). It is therefore essential to develop the capacity of partners in community-based organizations to enable them to implement basic systems that ensure adequate data management and quality control. To assess capacity of partner or site MERL systems, we start by conducting an office visit where we meet with the organization to explain the concept of data quality and to obtain general information from the organization on its data procedures. During this visit, we complete an assessment using **Part 2 of the 'Service Delivery Site' tab of the RDQA Excel Tool**. This visit happens during *[add when: during the pre-award process or as soon as we have signed an agreement?]*

D. Reports and Action Plans

- D1: Data Quality Assessment Dashboard Report
- D2: Action Plans

D1: Data Quality Assessment Dashboard Report

Plans for corrective action are developed after routine data review or RDQA and should include measures taken to address the risks to data quality and strengthen the MERL systems. These plans are updated *[enter frequency of updating, e.g. annually, bi-annually]* to ensure that they are relevant to the prevailing situation with regards to data quality.

The plan for corrective action includes the following:

- What we found during our assessments.
- Any trends across sites/sub-partners/regions.
- What remedial corrective actions will be required.
- When/how often we plan to go back to see if the corrective actions have been implemented.
- When we will recheck the actual data to make sure it's fixed and is now accurate.
- Who the responsible person is for following up on the Action Plan, both from the Pact side and Partner side.

[insert summary dashboards here, along with your analysis]

D2: Action Plans

After examination of the findings displayed in the dashboard of the Excel tool, we create an Action Plan for each level as follows:

RDQA Final Action Plan

Program/project				
Date of RDQA				
Date of Proposed Follow-up				
Description of Action Point	Person Responsible	Time Line	Technical assistance needs	Follow-up date and comments

References

International Statistics Institute. 1985. Declaration on professional ethics. The Hague, Netherlands: International Statistical Institute. <http://www.isi-web.org/about-isi/professional-ethics/43-about/about/296-declarationprofessionalethics-2010uk>

MEASURE Evaluation. *Tools for data demand and use*. Chapel Hill, North Carolina: Carolina Population Center. <http://www.cpc.unc.edu/measure/tools/data-demand-use/data-demand-and-use-strategies-and-tools.html>.

MEASURE Evaluation. 2011. *Tools for data demand and use in the health sector: Framework for linking data with action*. Chapel Hill, North Carolina. Carolina Population Center. <http://www.cpc.unc.edu/measure/publications/ms-11-46-b> (accessed August 29, 2013).

McIntire, Todd. 2002. The administrator's guide to data quality. *Technology and Learning* 22 (June): 18–33. <http://www.jcu.edu/education/dshutkin/ed585/DataDrivenDecisions.pdf>.

Plaza, Joyce, and Ruth Fischbach. 2012. Privacy and confidentiality. *Current Issues in Research Ethics*. New York: Columbia Center for New Media Teaching & Learning. http://ccnmtl.columbia.edu/projects/cire/pac/foundation/index.html#1_1 (accessed December 11, 2012).

The Global Fund to Fight Aids, Tuberculosis and Malaria; Office of the Global AIDS Coordinator, PEPFAR; USAID; WHO; UNAIDS; MEASURE Evaluation. 2008. *Routine data quality assessment tool: Guidelines for implementation* [draft]. <http://www.cpc.unc.edu/measure/tools/monitoring-evaluation-systems/data-quality-assurance-tools/RDQA%20Guidelines-Draft%207.30.08.pdf>

MEASURE Evaluation. 2011. *Tools for data demand and use in the health sector: Assessment of data use constraints*. Chapel Hill, North Carolina: MEASURE Evaluation. <http://www.cpc.unc.edu/measure/publications/ms-11-46-a> (accessed August 29, 2013).