Service Availability and Readiness Assessment (SARA)

An annual monitoring system for service delivery



© World Health Organization 2015

All rights reserved. Publications of the World Health Organization are available on the WHO website (www.who.int) or can be purchased from WHO Press, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland (tel.: +41 22 791 3264; fax: +41 22 791 4857; e-mail: bookorders@who.int).

Requests for permission to reproduce or translate WHO publications –whether for sale or for non-commercial distribution– should be addressed to WHO Press through the WHO website (www.who.int/about/licensing/copyright_form/en/index.html).

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the World Health Organization in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by the World Health Organization to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall the World Health Organization be liable for damages arising from its use.

Cover photo credit: WHO/Evelyn Hockstein

Service Availability and Readiness Assessment (SARA)

An annual monitoring system for service delivery

Reference Manual

Version 2.2 Revised July 2015



Acknowledgements

The service availability and readiness assessment (SARA) methodology was developed through a joint World Health Organization (WHO) – United States Agency for International Development (USAID) collaboration. The methodology builds upon previous and current approaches designed to assess service delivery including the service availability mapping (SAM) tool developed by WHO, and the service provision assessment (SPA) tool developed by ICF International under the USAID-funded MEASURE DHS project (monitoring and evaluation to assess and use results, demographic and health surveys) project, among others. It draws on best practices and lessons learned from the many countries that have implemented health facility assessments as well as guidelines and standards developed by WHO technical programmes and the work of the International Health Facility Assessment Network (IHFAN).

Particular thanks are extended to all those who contributed to the development of the service readiness indicators, indices, and questionnaires during the workshop on "Strengthening Monitoring of Health Services Readiness" held in Geneva, 22–23 September 2010.

Many thanks to The Norwegian Agency for Development Cooperation (Norad) whom has supported Statistics Norway to take part in the development of the SARA tools. The support has contributed to the development and implementation of a new electronic questionnaire in CSPro and data verification guidelines.

A special thanks to the Medicines Information and Evidence for Policy unit at WHO for their contribution to the SARA training materials and to the Unidad de Calidad y Seguridad de la Atención Médica-Hospital General de México for their contribution of photographs to the SARA data collectors' guide.

Project Management Group

The SARA methodology and tool were developed under the direction and management of Kathy O'Neill and Ashley Sheffel with valuable inputs from Ties Boerma and Marina Takane.

Project Advisory Group

Carla AbouZahr, Maru Aregawi Weldedawit, Sisay Betizazu, Paulus Bloem, Krishna Bose, Maurice Bucagu, Alexandra Cameron, Daniel Chemtob, Meena Cherian, Richard Cibulskis, Mario Dal Poz, Sergey Eremin, Jesus Maria Garcia Calleja, Sandra Gove, Neeru Gupta, Teena Kunjumen, Thierry Lambrechts, Richard Laing, Blerta Maliqi, Shanthi Mendis, Claire Preaud, Andrew Ramsay, Leanne Riley, Cathy Roth, Willy Urassa, Adriana Velasquez Berumen, Junping Yu, Nevio Zagaria, and Evgeny Zheleznyakov.

Table of contents

Ackn	Acknowledgements2					
Abbr	Abbreviations4					
CHAF	TER 1 OVERVIEW	.5				
1.1	Background	.7				
1.2 1.3	Pre-survey preparation	.9 16				
1.4	Planning the survey	22				
1.5	Training field supervisors, data collectors and data entry personnel2	25				
1.6	Preparing for data collection in the field	30				
1./ 1 8	Data collection in the field	36 10				
1.9	Data analysis	46				
1.10	Data archiving	55				
Refer	ences	53				
CHAF	TER 2 CORE INSTRUMENT6	55				
CHAF	TER 3 INDICATORS INDEX	25				
3.1	Indicators ID numbers	27				
3.2	SARA general service availability indicators	27				
3.3	SARA general service readiness indicators	31				
3.4	SAKA Service specific availability and readiness indicators	50				

Abbreviations

AIDS	acquired immunodeficiency syndrome
ALT	alanine aminotransferase
CBR	crude birth rate
CSV	comma-separated values
DBS	dried blood spot
DCMI	Dublin Core Metadata Initiative
DDI	Data Documentation Initiative
DQRC	Data quality report card
DV	Data verification
EDC	electronic data collection device
FBO	faith-based organization
GIS	geographical information system
GPS	global positioning system
HIV	human immunodeficiency virus
HMIS	health management information system
HRIS	human resources information system
ID	identification
IHFAN	International Health Facility Assessment Network
IHP+	International Health Partnership and related initiatives
IHSN	International Household Survey Network
M&E	monitoring and evaluation
MDG	Millennium Development Goal
MFL	master facility list
MNCH	maternal, newborn and child health
МоН	ministry of health
NADA	national data archive
NGO	nongovernmental organization
OECD	Organisation for Economic Co-operation and Development
PMTCT	prevention of mother-to-child transmission (of HIV)
RDT	rapid diagnostic test
SAM	service availability mapping
SARA	service availability and readiness assessment
SPA	service provision assessment
UNAIDS	Joint United Nations Programme on HIV/AIDS
UNDP	United Nations Development Programme
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development
WHO	World Health Organization
XML	extensible markup language

1. Overview

1.1 Background

1.1.1 Why measure service availability and readiness?

Sound information on the supply and quality of health services is necessary for health systems management, monitoring and evaluation. Efforts to achieve the Millennium Development Goals (MDGs) and to scale up interventions for HIV/AIDS, malaria, safe motherhood and child health through global health partnerships, have drawn attention to the need for strong country monitoring of health services, covering the public, private-for-profit and private not-for-profit sectors, and their readiness to deliver key interventions.

With the increased demand for accountability and the need to demonstrate results at country and global levels, information is needed to track how health systems respond to increased inputs and improved processes over time, and the impact such inputs and processes have on improved health outcomes and better health status. However, despite heightened investments in health systems, few countries have up-to-date information on the availability of health systems that covers both the public and private sectors. Fewer still have accurate, up-to-date information required to assess and monitor the "readiness" of health facilities to provide quality services.

Ensuring access to quality health services is one of the main functions of a health system. Service access includes different components: availability, which refers to the physical presence or reach of the facilities; affordability, which refers to the ability of the client to pay for the services; and acceptability, which refers to the sociocultural dimension.

The quality of services is yet another dimension. A prerequisite to service quality is service readiness, i.e. the health facilities should have the capacity to deliver the services offered. This capacity includes the presence of trained staff, guidelines, infrastructure, equipment, medicines and diagnostic tests. Service availability and readiness are prerequisites to quality services, but do not guarantee the delivery of quality services.

1.1.2 The global and country context

Building upon principles derived from the Paris Declaration on Aid Effectiveness and the International Health Partnership and related initiatives (IHP+), global partners and countries have developed a general framework for the monitoring and evaluation (M&E) of health system strengthening (1). This framework centres on country health strategies and related M&E processes such as annual health sector reviews, and at its core is the strengthening of a common monitoring and review platform to improve the availability, quality and use of data to inform health sector review processes and global monitoring (2).

Within this context, WHO has been working with USAID, MEASURE Evaluation, MEASURE DHS, ICF International, and other country and global partners to develop tools to fill critical data gaps in measuring and tracking progress in health systems strengthening. Service availability and readiness assessment (SARA) is one tool available to fill data gaps on service delivery.

SARA relies on a rapid data collection and analysis methodology, and can be combined with a record review to assess data quality of the facility reporting system. Ideally, SARA is conducted approximately three to five months prior to a health sector review to allow for the results to feed into the health sector review process.

1.1.3 Related surveys and initiatives

The service availability and readiness assessment (SARA) effort builds on previous and current approaches designed to assess health facility service delivery including the service availability mapping (SAM) tool developed by WHO (*3*), and the service provision assessment (SPA) tool developed by ICF International under the USAID-funded MEASURE DHS project (*4*).

The SARA methodology takes into account best practices and lessons learned from the many countries that have implemented health facility assessments of service availability and readiness. It also draws heavily on the work of the International Health Facility Assessment Network (IHFAN) and on experiences from programmeand service-specific facility assessment work.

The training materials for SARA draw on best practices and materials developed for survey methods such as the SPA and the WHO/Health Action International (HAI) methodology for measuring medicine prices, availability, affordability and price components (5).

1.2 Survey overview

1.2.1 Survey objectives

SARA is designed as a systematic survey to assess health facility service delivery. The objective of the survey is to generate reliable and regular information on service delivery including service availability, such as the availability of key human and infrastructure resources, and on the readiness of health facilities to provide basic health-care interventions relating to family planning, child health services, basic and comprehensive obstetric care, HIV/AIDS, tuberculosis, malaria and noncommunicable diseases.

The SARA survey generates a set of tracer indicators of service availability and readiness that can be used to:

- detect change and measure progress in health system strengthening over time;
- plan and monitor the scale-up of interventions that are key to achieving the MDGs, such as implementing interventions to reduce child and maternal mortality, HIV/AIDS, tuberculosis and malaria, and to respond to the increasing burden of noncommunicable diseases;
- generate the evidence base to feed into country annual health reviews, to better inform the development of annual operational plans and to guide more effective country and partner investments;
- support national planners in planning and managing health systems (e.g. assessing equitable and appropriate distribution of services, human resources and availability of medicines and supplies).

Key outputs from SARA form the basis for national and subnational monitoring systems of general service availability and readiness, and service-specific readiness (maternal and child health, HIV/AIDS, tuberculosis, malaria, noncommunicable diseases, surgical care, etc.). SARA products include a regularly updated national database of public and private facilities, and an analytical report of core indicators to assess and monitor availability of health services and readiness to provide services.

QUESTIONS ANSWERED BY SERVICE AVAILABILITY AND READINESS ASSESSMENT (SARA)

- What is the availability of basic packages of essential health services offered by public and private health facilities?
- Is there an adequate level of qualified staff?
- Are resources and support systems available to assure a certain quality of services?
- How well prepared are facilities to provide high-priority services such as reproductive health services, maternal and child health services, and infectious disease diagnosis and treatment (e.g. HIV, sexually transmitted infections, tuberculosis and malaria)?
- Are facilities ready to respond to the increasing burden of noncommunicable diseases?
- What are the strengths and weaknesses in the delivery of key services at health-care facilities?

1.2.2 Key topics, indicators and indices

The SARA survey is designed to generate a set of core indicators on key inputs and outputs of the health system, which can be used to measure progress in health system strengthening over time (6). Tracer indicators aim to provide objective information about whether or not a facility meets the required conditions to support provision of basic or specific services with a consistent level of quality and quantity. Summary or composite indicators, also called indices, can be used to summarize and communicate information about multiple indicators and domains of indicators. Indices can be used for general and service-specific availability and readiness.

There are three main focus areas of SARA.

- I. Service availability refers to the physical presence of the delivery of services and encompasses health infrastructure, core health personnel and aspects of service utilization. This does not include more complex dimensions such as geographical barriers, travel time and user behaviour, which require more complex input data. Service availability is described by an index using the three areas of tracer indicators (see Table 1.2.1). This is made possible by expressing the indicators as a percentage score compared with a target or benchmark, then taking the mean of the area scores.
- II. General service readiness refers to the overall capacity of health facilities to provide general health services. Readiness is defined as the availability of components required to provide services, such as basic amenities, basic equipment, standard precautions for infection prevention, diagnostic capacity and essential medicines. General service readiness is described by an index using the five general service readiness domains (see Table 1.2.1). A score is generated per domain based on the number of domain elements present, then an overall general readiness score is calculated based on the mean of the five domains.
- III. **Service-specific readiness** refers to the ability of health facilities to offer a specific service, and the capacity to provide that service measured through consideration of tracer items that include trained staff, guidelines, equipment, diagnostic capacity, and medicines and commodities.

Domain	Tracer indicators, items or services
I. Service availability	
1. Health infrastructure	 Number of health facilities per 10 000 population Number of inpatient beds per 10 000 population Number of maternity beds per 1000 pregnant women
2. Health workforce	Number of health workers per 10 000 population
3. Service utilization	 Outpatient visits per capita per year Hospital discharges per 100 population per year

 TABLE 1.2.1: SUMMARY OF TRACER INDICATORS, ITEMS AND SERVICES FOR SERVICE AVAILABILITY AND SERVICE

 READINESS

Domain	Tracer indicators, items or services			
II. General service readiness				
1. Basic amenities	Mean availability of seven basic amenities items (%): power, improved water source, room with privacy, adequate sanitation facilities, communication equipment, access to computer with Internet, emergency transportation			
2. Basic equipment	Mean availability of six basic equipment items (%): adult scale, child scale, thermometer, stethoscope, blood pressure apparatus, light source			
3. Standard precautions for infection prevention	Mean availability of 9 standard precautions items (%): safe final disposal of sharps, safe final disposal of infectious wastes, appropriate storage of sharps waste, appropriate storage of infectious waste, disinfectant, single-use disposable/auto-disable syringes, soap and running water or alcohol-based hand rub, latex gloves and guidelines for standard precautions			
4. Diagnostic capacity	Mean availability of 8 laboratory tests available on-site and with appropriate equipment (%): haemoglobin, blood glucose, malaria diagnostic capacity, urine dipstick for protein, urine dipstick for glucose, HIV diagnostic capacity, syphilis RDT and urine pregnancy test			
5. Essential medicines	Mean availability of 25 essential medicines (%):Amlodipine tablet or alternative calcium channel blocker, amoxicillin (syrup/suspension or dispersible tablets), amoxicillin tablet, ampicillin powder for injection, aspirin (capsules/tablets), beclometasone inhaler, beta blocker (e.g.bisoprolol, metaprolol, carvedilol, atenolol), carbamazepine tablet, ceftriaxone injection, diazepam injection, enalapril tablet or alternative ACE inhibitor (e.g. lisonopril, Ramipril, perindopril), fluoxetine tablet, gentamicin injection, glibenclamide tablet, haloperidol tablet, insulin regular injection, magnesium sulfate injectable, metformin tablet, omeprazole tablet or alternative (e.g. pantoprazole, rabeprazole), oral rehydration solution, oxytocin injection, salbutamol inhaler, simvastatin tablet or other statin (e.g. atorvastatin, pravastatin, fluvastatin), thiazide (e.g. hydrochlorothiazide) and zinc sulphate (tablet or syrup).			
III. Service-specific readiness				
For each service, the readiness score is computed as the mean availability of service-specific tracer items in four domains: staff and training, equipment, diagnostics, and medicines and commodities	 Family planning Antenatal care Basic obstetric care Comprehensive obstetric and neonatal care Child health immunization Child health preventative and curative care Adolescent health services Lifesaving commodities for women and children Malaria diagnosis or treatment Tuberculosis services HIV counselling and testing HIV/AIDS care and support services Antiretroviral prescription and client management Prevention of mother-to-child transmission (PMTCT) of HIV Sexually transmitted infections diagnosis or treatment Noncommunicable diseases diagnosis or management: diabetes, cardiovascular disease, chronic respiratory disease and cervical cancer screening Basic and comprehensive surgical care Blood transfusion Laboratory capacity 			

1.2.3 Core instrument

The basic approach to SARA is to collect data that are comparable both across countries and within countries (i.e. across regions and/or districts). To achieve this, a standard core questionnaire has been developed. The core questionnaire was pretested in a variety of settings in two countries. The first pretest occurred in Sierra Leone in April, 2011. A second pretest occurred in Kenya in June, 2011. This second pretest was part of a larger pretest of the revised MEASURE DHS Service Provision Assessment (SPA) questionnaire, which includes all core SARA questions as they are embedded in the SPA Inventory questionnaire. Following the pilot test experience, adjustments were made to the questionnaire to account for the information gained, resulting in the standard core questionnaire.

Typically, a country adopts the core questionnaire with adaptations to certain elements such as:

- types of facilities
- managing authority of facilities
- national guidelines for services
- staffing categories
- national policies for medicines (e.g. for tuberculosis, HIV/AIDS).

The questionnaire does not attempt to measure the quality of services or resources, but it can be used in conjunction with additional modules such as management assessment or quality of care.

1.2.4 Survey design methodology

The SARA survey requires visits to health facilities with data collection based on key informant interviews and observation of key items. The survey can either be carried out as a sample or a census; the choice between these methodologies will depend on a number of elements including the country's resources, the objectives of the survey and the availability of a master facility list (MFL). For example, if the objective of the survey is to have nationally representative estimates, a sample survey would be appropriate. However, if the objective is to have district estimates, the sampling methodology must be adjusted to either a larger sample or in some cases a full census.

Service availability

The recommended data source for information on service availability is a national master facility list (MFL) of all public and private facilities (7). A facility census is usually required to establish and maintain a national MFL. A facility census aims to cover **ALL** public and private health facilities in a country. The census is designed to form the basis for a national and subnational monitoring system of service delivery, which can be supplemented by quality ascertainment through facility surveys and further in-depth assessments. A census is the recommended methodology for forming the baseline of service availability and readiness data. Service availability data should be updated annually through routine, facility-based reporting and validated approximately every five years through a facility census.

Service readiness

The recommended design methodology for measuring service readiness is a sample survey. Sampling is done in a systematic way to ensure that the findings are representative of the country and region/district in which the survey is being conducted. Drawing a random sample of health facilities is much more complicated if the country does not have a comprehensive and up-to-date MFL. Therefore, it is highly recommended to invest in establishing a MFL that includes all public and private facilities. In cases where a national list of facilities is not available or up-to-date, the service readiness survey can be carried out at the same time as the facility census for service availability.

Master facility list (MFL)

Regardless of the method selected, a complete MFL is required. Therefore, it is highly recommended to invest in establishing a MFL that includes all public and private health facilities. In many countries there are already lists of public facilities and sometimes also nongovernmental facilities. However, private facilities are often excluded or only partially included in these lists. WHO and partners have developed a guide to support countries in creating a MFL. Please refer to the document *Creating a master facility list (7)* for more information on best practices in establishing a MFL.

Data quality assessment

The service availability and readiness assessment can be used for to assess data quality of the routine system by comparing results with aggregated routine health information data at district, provincial and national level. In addition, the service readiness assessment can be combined with a record review for data verification purposes, to ascertain the completeness and quality of facility reporting. The data quality review (DQR) (*8*) can be used to verify the quality of routinely reported data for some key coverage indicators, quantifying problems of data completeness accuracy and external consistency.

1.2.5 Timeline of implementation

Service availability and readiness assessments should be planned on a yearly or biennial basis to coincide with and feed into national health planning cycles. Sample surveys should be organized every year about three to five months in advance of the annual review. The national MFL should be used to provide the sampling frame (see Figure 1.2.1).



FIGURE 1.2.1: TIMELINE OF IMPLEMENTATION FOR SERVICE AVAILABILITY AND READINESS ASSESSMENT

The time needed to complete a service availability and readiness assessment depends on the size of the country and whether or not there is a need for a full facility census. From the initial country-adaptation of the assessment tool to the dissemination of data and production of country reports, the entire process generally takes from three to six months.

Table 1.2.2 below provides an overview of the survey steps and the activities to be undertaken at each step.

Steps	Survey activities
1. Survey planning and preparation	 Establish a survey coordinating group of country stakeholders to oversee and facilitate the objectives, scope, design, implementation and analysis Obtain a list of all health facility sites (public, private, nongovernmental organizations (NGOs) and faith-based organizations (FBOs)), including country facility registry codes Determine appropriate design methodology (census or sample), develop an implementation plan and budget, and secure funding Adapt questionnaires to meet country-specific needs Recruit survey personnel (survey manager, field supervisors, data collectors, data entry/processing personnel, data analysts) Prepare a survey schedule Identify the survey sites (sampling frame). Select the sample size and sample of health facilities (if sampling methodology is chosen) Procure logistics including equipment and transport, taking into consideration the number of sites to be visited, the number of data collection teams, drivers, vehicles, petrol, etc. Plan and conduct training courses for interviewers and field supervisors Pilot test the survey in a selected number of health facilities, evaluate results and make amendments if necessary
2. Data collection in the field	 Plan the data collection visits (prepare a letter of introduction, contact each site, prepare a schedule of visits) Prepare materials and tools for data collectors Arrange for transport and regular communications during fieldwork Assemble materials necessary for local data collection Confirm appointments with health facilities Visit health facilities and collect SARA data in teams (usually two interviewers and a driver) At the end of the interview, check questionnaire and resolve missing/unreliable information Return completed forms and/or transfer electronic files to field supervisor at the conclusion of each day Return forms (paper and/or electronic) to survey manager when data collection is complete
3. Data entry, analysis and interpretation	 Enter data using the CSPro application Edit, validate and clean data set, checking for consistency and accuracy Export the data set for analysis (SARA indicators) Conduct analyses of SARA data using the standard core indicators (SARA automated tool for results graphs and tables) as well as any country-specific indicators of interest
4. Results dissemination	 Meet with survey coordinating group to analyze and interpret survey results and to finalize recommendations Prepare the final report Plan and implement dissemination activities. The results should be used to support annual health reviews and feed into the M&E platform for the national health plan Document and archive the survey using metadata standards

1.2.6 Roles and responsibilities

The survey is usually undertaken under the overall leadership of the **Ministry of Health**. The following section briefly outlines the roles and responsibilities of the key parties involved in the implementation of SARA and data quality activities.

Ministry of Health (MoH): Will have overall responsibility for the coordination of this process. Will coordinate and provide support to get permission to conduct data collection activities, help with the coordination of analysis and results dissemination meetings by inviting all the appropriate governmental departments, key non-governmental and development partners. Will also promote the use of this data for policy and planning.

Survey Coordinating Group: The Coordinating Group, led by Ministry of Health should include national institutes and other key stakeholders in the health services sector. This core group, will provide leadership and oversight throughout the whole process from questionnaire to dissemination of results.

Implementation agency: Will be responsible for conducting field data collection for SARA and the data verification component of the Data Quality Review. Details of the composition of the implementation agency team is given in a separate document.

<u>Agency providing quality assurance and technical support</u>: It is recommended that an independent party be involved in the implementation process. This support can be provided by a separate national institute or independent consultant. He/she will be responsible for providing support to the implementation team on planning and implementing SARA; provide a quality assurance role to ensure due processes are followed during training, data collection, cleaning and analyses stages (including validation visits in 10% of the facilities); to provide assistance and oversight to the implementing team on the production of the SARA and data quality assessment report.

1.3 Pre-survey preparation

1.3.1 Establishing a survey coordinating group

Bringing partners together and mobilizing them around the survey is a critical first step towards successful implementation. One of the first activities is to identify and establish a group of core stakeholders at country level to oversee, coordinate and facilitate the planning, implementation and follow-up of the facility assessment process. In general, partners include those groups, individuals, and/or organizations that are carrying out or planning similar efforts as well as those for whom the outputs of the health facility assessment will be of interest. These often include:

- ministries of health (as well as national institutes of statistics, geographical information system (GIS) units, health management information systems (HMIS) units, health services and other public research institutions);
- universities and other academic institutions involved in research;
- NGOs and other organizations involved in data collection;
- United Nations health-related organizations present in the country (e.g. WHO, UNICEF, UNDP, UNAIDS);
- international funders active in the country (i.e. the Global Fund to Fight AIDS, Tuberculosis and Malaria, government agencies for international development).

The role of the survey coordinating group should include:

- clarifying the objectives of the survey;
- supporting the survey manager in planning, preparing and conducting the study, and identifying important policy issues that should inform the survey protocol;
- advising on any matters that arise during survey preparation, fieldwork and data analysis;
- assisting in interpreting data and developing policy recommendations;
- promoting the findings of the survey and advocating for appropriate policy recommendations.

It is important to hold regular meetings with the survey coordinating group throughout the survey process. At least one meeting should be held to support the planning and preparation of the SARA survey, and one meeting should be held post-survey for interpreting survey results and developing recommendations. A second post-survey meeting may be beneficial to discuss the results and their policy implications, consolidate all survey results and finalize recommendations.

1.3.2 Identifying entities for survey implementation and quality assurance

Once the coordinating group has been established, it is important to define who will be in charge of the survey field implementation. It is recommended to identify and work with a national institute (e.g. National Statistical Office, School of Public Health, etc.) or an entity used to conducting such field assessments. The selection is done in agreement with the Ministry of Health.

The institute in charge of the survey implementation works closely with the coordinating group for the preparation, implementation and results dissemination of the survey. It is also recommended that a third party (different from the implementing agency – consultant, regional institute, etc.) ensures the quality of the survey (including organizing visits in 10% of the sites). This entity can also provide technical backstopping as required. It works closely with the coordinating group and implementation institute throughout the process, ensuring that survey procedures are followed properly and as per the defined methodology.

1.3.3 Compiling a master facility list (MFL)

Before beginning a health facility assessment, a situation analysis assessing the availability of health facility information must be carried out. An important prerequisite for conducting a SARA survey is the existence of a MFL. The analysis should therefore aim to ascertain the existence and reliability of an official MFL.

Regardless of the survey methodology (census or sample), a complete master list of facilities is required. Therefore, it is highly recommended to invest in establishing a MFL that includes public, private for-profit, and NGO facilities. In many countries there are already lists of public facilities and sometimes also nongovernmental facilities. However, the private facilities are often excluded or only partially included in these lists.

Before a health facility assessment can be implemented, **ALL** health facilities in a country must be identified and a health facility list created. This list must include health facilities in all sectors including the public sector, the private sector, FBOs and NGOs. In some countries, a MFL may be available containing all the required information. In most cases however, this information is not readily available and must be compiled. The ministry of health (MoH) generally maintains information on public health facilities and can serve as the basis for the MFL. Other contacts will need to be identified to retrieve information on private, FBO, NGO and other facilities.

All available health facility listings have to be reconciled to identify a single, comprehensive list, with each facility assigned a unique identification (ID) code. Facilities should be classified by level of service provision (from hospital at the highest level through clinic at the lowest level) and by ownership (MoH, mission, NGO or private). Locational information should be included in the MFL when available. The geographical coordinate collection method should also be recorded (i.e. global positioning system (GPS) remote device, digital place names, gazetteers, etc.).

A key component of the MFL is the unique ID code assigned to each facility. A set of data must be gathered with the specific purpose of uniquely identifying each survey site. In database terminology this set of identifier data is referred to as a "primary key" or a "unique key": a code uniquely identifying a row or column of a database. Without specific ID attached to each survey site, there is a risk of duplicate data collection. In addition to greatly lessening the risk of data duplication, site ID fields allow for cross-survey comparisons as well as comparisons over time.

Please refer to the document *Creating a master facility list* (7) for more information on best practices in establishing a MFL.

1.3.4 Designing a methodology and implementation plan

Design methodology

There are two potential design methodologies for the SARA survey:

- a facility census (i.e. assessment of all health facilities)
- a sample survey (i.e. a representative sample of facilities).

Service availability requires a denominator that includes all public and private health facilities in the country (i.e. a census). Service readiness can be measured through a representative sample of facilities.

Facility census

The recommended data source for information on service availability is a national MFL of all public and private facilities. A facility census is usually required to establish and maintain the MFL. Service availability data should be updated annually through routine, facility-based reporting, and data should be validated approximately every five years through a facility census.

A facility census aims to cover ALL public and private health facilities in a country. The census is designed to form the basis for a national and subnational monitoring system of service delivery, which can be supplemented by quality ascertainment through facility surveys and further in-depth assessments. A census is the recommended methodology for forming the baseline of service availability and readiness data.

Sample survey

The recommended design methodology for measuring service readiness is a sample survey. Sampling is done in a systematic way to ensure that the findings are representative of the country or state/province in which the survey is being conducted. Drawing a random sample of health facilities will be much more complicated if the country does not have a comprehensive and up-to-date MFL. Therefore, it is highly recommended to invest in establishing a MFL that includes public, private for-profit, and nongovernmental facilities. If a fairly complete master list of facilities already exists, a sampling approach can be used.

Implementation plan and budget

An implementation plan should be drafted based on the objectives of the survey and the results of the situation analysis of health facility information. The implementation plan serves as a comprehensive outline of the operational plan for implementing a SARA survey and is key to ensuring the success of the survey. The plan must lay out the reason for carrying out the survey, how the survey will be executed and how to oversee the survey to ensure that it will be completed on time and within budget. The objectives of the survey will help to determine the design methodology, which will in turn drive much of the operational plan and budget for the survey. When designing the budget, it is essential to ensure that the following items are accounted for.

Financial and human resources

- human resources field supervisors data entry personnel survey manager TA/QA entity data collectors data analysts training training venue transport expenses related to pilot testing daily allowance and materials accommodation data collection and validation daily field allowance materials (paper, communication (e.g. and accommodation pens, etc.) telephone charges) for data collectors photocopying transport data cleaning, processing and analysis meetings of the survey coordinating group
 - report production and dissemination
 - advocacy and communications
- overheads
- contingency.

Technical resources

- mobile electronic data collection devices (EDCs) e.g. personal digital assistants (PDAs), tablet computers or laptop computers: one for each data collection team
- GPS devices (if the facility coordinates need to be taken): one for each data collection team
- batteries for GPS devices
- memory cards for EDCs
- computer(s) for data entry
- data entry application (CSPro¹ recommended)
- data analysis programme.

Once a comprehensive budget has been developed, funding should be secured to cover all survey costs (a standard budget template is available in the SARA Implementation Guide—Chapter 1: Planning).

1.3.5 Adapting the SARA instrument to country-specific needs

A standard core questionnaire for measuring service availability and readiness is available. However, the questionnaire must be adapted for country use to reflect the needs of each country and specificities of each health-care system. When adapting the health facility questionnaire, consideration should be given to how changes will affect data collection, and adjustments should be made to ascertain that definitions are specific enough to assure comparability across the country and within districts. Training, data collection and analysis are carried out, even in larger countries, within one month, and adding more to the tool will make it slower and could create problems at the analysis stage if not carefully considered. SARA is not intended to provide comprehensive data on all aspects of health system functioning. Rather, it focuses on key "tracer" elements that are critical to programmes that are scaling up or that are indicative of the essential health system underpinnings or "readiness" to do so. This should be kept in mind while adapting the questionnaire and adding additional modules or questions.

The following areas of the SARA tool must always be adapted to the country context:

- types of facilities
- managing authority of facilities
- national guidelines for services
- staffing categories
- tuberculosis medicines
- HIV/AIDS medicines
- Routine immunization
- other country-specific medicines.

The questionnaire can be implemented as either a paper questionnaire or an electronic questionnaire.

Paper questionnaire: any changes should be made according to the country adaptation process prepared for the survey training.

Electronic questionnaire: once a mobile EDC has been selected, the appropriate software can be chosen. This software generally comprises a desktop forms designer and database, a synchronization conduit and the handheld forms application. Once the software is uploaded, the survey form can be designed on a desktop

¹ For information about the Census and Survey Processing System (CSPro), including free download, visit: <u>http://www.census.gov/population/international/software/cspro/</u>

computer and then synchronized with the handheld device. For the SARA survey, the recommended software for electronic questionnaires is CSPro.

1.3.6 Recruiting survey personnel

The SARA survey will require involvement of the following personnel:

- survey manager
- field supervisors
- data collectors
- data entry personnel
- data analysts.

Survey manager

The survey manager plans and coordinates the survey at the central (national) level. This includes planning the survey's technical and logistical aspects, recruiting and training survey personnel, supervising data collection and data entry, conducting data quality assurance and data analysis, interpreting results and preparing a survey report.

Wherever possible, the survey manager should have experience in conducting surveys and should be very familiar with the health-care system. The survey manager should be familiar with basic statistics and interpreting data. Successful communication of the survey results also requires an understanding of the policy-making process and different advocacy strategies. Where the survey manager does not possess all of these qualities, he or she should select the survey coordinating group members to ensure that the survey management team includes the necessary health, surveying, statistics, policy and advocacy skills.

Field supervisors

Field supervisors are responsible for overseeing all aspects of data collection in the survey area(s) for which they are responsible. In a small country or in a survey that is conducted in a single region of a country, it may be possible for all fieldwork to be undertaken by a single team. Experience has shown that in larger-scale studies, however, it is advisable to designate a field supervisor in each of the geographical areas that will be surveyed.

Field supervisors have a crucial role to play in ensuring data quality and consistency. They should be experienced in data collection and be familiar with health terminology. They are also instrumental in gaining access to facilities; if any field supervisor is unfamiliar with their designated area, a local contact may be needed to assist. Field supervisors may also be responsible for choosing local data collectors when they are not sent from the central level.

Data collectors

Data collectors are responsible for visiting health facilities and collecting SARA data with a high degree of accuracy. The survey methodology has been designed to minimize as far as possible the need for a high level of technical expertise.

However, data collectors should, wherever possible, have the following qualifications, skills and capabilities:

- a health qualification (nurse, midwife or medical student) and familiarity with the organization and functioning of health facilities;
- some understanding of the principles of sample surveys, ideally with some previous experience in conducting surveys;
- an appreciation of the logistics requirements for carrying out field studies;

- post-secondary school education as a minimum;
- familiarity with the locality and local language or dialect.

Data collection requires an aptitude for concentration and attention to detail. The best data collectors combine the discipline of collecting data in a standardized way with the ability to identify unusual situations that require advice from the field supervisor or survey manager. Data collectors must be available to work full time for the duration of the fieldwork. They should be willing to work extended hours if necessary and be able to stay away from their homes for extended periods of time.

The number of data collectors required depends on the sample size of the survey. Data collectors should work in pairs. Each visit to a health facility is likely to require about two hours plus transport time. In practice, this means that a team of two data collectors can survey two to four facilities per day. The number of data collectors will also depend on the budget available, the locations of the survey areas, the travel conditions and the number of health facilities to be surveyed. It is better to have a smaller number of better qualified data collectors than to have a large team where some data collectors lack the necessary skills.

Data entry and data processing personnel

Accurate data entry is vital to ensure the reliability of the results. Two data processing personnel with experience in using the selected data entry software are required: one to enter the data, and the other to reenter the same data to check that the entries are correct. If data are being entered from paper questionnaires, double entry is essential to ensuring the accuracy of the data entry process. If data are collected both electronically and on paper, then the first instance of data entry has already occurred during the electronic data collection and the data entry personnel would only be responsible for the second entry of data for validation purposes. In some cases, it may be possible to use the same personnel for both data collection and data entry, provided they have the necessary expertise to undertake both functions.

Data analysts

The primary tasks of the data analyst(s) are to inspect, clean, transform, analyse and visualize data with the goal of highlighting useful information, suggesting conclusions and supporting decision-making. It is vital that the data analyst has an advanced knowledge of the chosen analysis software for the SARA survey. A working knowledge of health service delivery and the specific country's health system is important for interpretation of the results and is required of at least one member of the data analysis team.

1.3.7 Preparing the survey schedule

The complete survey should generally take between three and six months to complete, including survey preparation, data collection, data entry, data analysis and report writing. Further time should be allotted for dissemination and follow-up activities. Given that the information gathered from SARA should be used to inform decision-making, it is important that data collection be conducted rapidly and the report generated as soon as possible once data collection is complete. This will ensure that the survey results are relevant and informative for decision-makers. A survey schedule should be developed and consulted regularly to ensure that activities are proceeding according to plan. This schedule should detail the amount of time allotted for each step in the survey process, and should serve as a timeline for all survey activities.

1.4 Planning the survey

1.4.1 Selecting the sample size and sample

Determining the sample size and selecting the sample for a facility survey is a complex subject, which will vary considerably from case to case depending on the desired precision and type of estimates, the number of facilities in the country as well as the specific objectives of the assessment. For example, a SARA conducted to produce national estimates will require a much smaller sample size than if district-level estimates are desired. In order to ensure that the sample is representative, it is best to consult with a sampling expert or a statistician to select an appropriate sampling methodology. For the SARA, the most common sampling strategy is Option 1 in the table below—a nationally representative sample obtained by taking a simple random sample of facilities within each stratum (facility type and managing authority) at the national level. The table below presents different sampling options that could be used to conduct a SARA based on the desired level of estimates:

Domains of estimation	Sampling method	Sample size (estimate) ²	Approximate cost
Option 1: National estimates only National estimates with disaggregation by facility type (3 levels) and managing authority (public/private)	Small country Stratification by facility type and managing authority, simple/systematic random sampling within each stratum with census or oversampling of hospitals (design effect = 1)	150 – 250 facilities	\$60K-100K
	Medium country Blend of list and area sampling: list sampling for large health facilities, and area sampling for small facilities (census of facilities in sampled area PSUs ³) (deff = 1.2)	250 – 500 facilities	\$100K-200K
Option 2: Subnational estimates Regional and national estimates with disaggregation by facility type (3 levels) and managing	Small country Stratification by region, facility type and managing authority, simple/systematic random sampling within each stratum, with census or oversampling of hospitals (deff = 1)	5 regions: 250 – 500 facilities 10 regions: 500 – 800 facilities	\$100K-130K \$130K-180K
authority (public/private)	<i>Medium/large country</i> Blend of list and area sampling: list sampling for large health facilities, and area sampling for small facilities (census of facilities in sampled area PSUs ²) (deff = 1.2)	<i>Medium country</i> 4 regions: 300 – 500 facilities <i>Large country</i> 4 regions: 400 – 800 facilities	\$120K-200K \$180K-360K
Option 3: Subnational estimates Regional estimates for a subset of regions, with disaggregation by facility type (3 levels) and managing authority (public/private) for selected regions; no national estimates	Large country Purposive sample of regions, simple/systematic random sample with oversampling of hospitals for each region	4 regions (150 facilities per region): 600 facilities	\$60-100K per region

² Sample size estimates assume a margin of error of 0.1 and 95% level of confidence

³ Administrative units that form the PSUs (Primary Sampling Units) for the area sample should contain approximately 1-5 health facilities each (communes, sub-counties, villages)

Domains of estimation	Sampling method	Sample size (estimate) ²	Approximate cost
Option 4: District sampleSmall, medium and large countriesDistrict estimates for sampled districts; national estimates if sufficiently many facilities are sampledList sampling of districts (two- 		Small country 300-500 facilities (10- 30 districts ⁴) Medium country 400-800 facilities (20+ districts) Large country 600-1000 facilities (30+ districts)	\$100K-200K \$160K-320K \$270K-470K
Option 5: Facility census All possible domains of estimation	Small, medium and large countries Census of all facilities		Very expensive

Small country: 50 – 100 hospitals, 1000 – 2000 health facilities total, 10 – 80 districts (e.g. Sierra Leone, Togo, Burkina Faso) *Medium country:* 100-500 hospitals, 2000 – 5000 health facilities total, 80 – 500 districts (e.g. Uganda, Tanzania) *Large country:* 500 – 1000 hospitals, 5000 – 10000 health facilities total, 500 – 1000 districts (e.g. DRC, Nigeria)

1.4.2 Procuring logistics

Planning for data collection requires consideration of the logistics needs for data collection teams as well as an assessment of the hardware and software needs for data collection. Equipment should be considered for a base camp as well as for fieldwork, and for operations as well as for training. The guiding principle that should be kept in mind when compiling equipment for the field is redundancy, i.e. to have backup components and a contingency plan in case equipment fails, breaks or is lost. All equipment should have one or more backups, depending on the equipment type and survey requirements. If feasible, paper forms and printing capabilities provide a viable contingency plan for the worst-case scenario of mobile device failure. Equipment requirements are also determined according to country-specific needs, as well as the availability of resources and budget.

Assigning facilities to teams

It is recommended to map all facilities in the survey sample to assist with logistics planning for the data collection. This map can be made either on paper or electronically. The map should include information such as roads, topography, basic geographical features, elevation and location of health facilities, which are useful in determining survey areas. Teams should be assigned to facilities based on the geographical distribution of the selected health facilities. Figure 1.4.1 gives an example of a map that would be useful for SARA logistics planning.

FIGURE **1.4.1** SERVICE AVAILABILITY AND READINESS ASSESSMENT (SARA) EXAMPLE MAP



⁴ Number of districts in sample depends on the number of facilities per district

Survey team requirements

The duration of the field survey depends on the availability of resources, the number of teams, the number of health facilities to be visited, and the size of the country and population.

As a general guide, data collection teams consist of two interviewers/data collectors plus a driver. On average, one team can cover at least two health facilities per day.

The estimated duration of the survey is calculated during the planning phase and is unique to the needs and resources available in the country. The following examples illustrate the planning that is required.

Example 1. A country consisting of 50 districts with on average 40 health facilities per district.

One team covers one district (40 health facilities) over 20 days (two facilities per day), and so 10 teams cover 10 districts over 20 days.

Therefore 10 teams will cover one country (50 districts) over 100 days (or three months).

Example 2. An urban area with an average of 200 health facilities.

One team covers 200 health facilities over 100 days (two facilities per day), or 10 teams cover 200 health facilities over 10 days.

For all surveys, logistics planning needs to take into account the following:

- car hire and fuel for the duration of the survey
- per diem for the driver(s)
- per diem for the data collectors.

Equipment requirements

Equipment requirements are also determined according to country-specific needs, as well as the availability of resources and budget.

1.5 Training field supervisors data collectors and data entry personnel

This chapter provides practical guidance on conducting a training workshop for field supervisors, data collectors and data entry personnel. Training is an important element of survey preparation because it helps to ensure the accuracy and reliability of the data gathering and data entry procedures. Consequently, this chapter also covers the issue of ensuring data quality. This chapter has been developed to assist survey managers in conducting training workshops for their survey personnel, regardless of whether they have attended any previous training.

1.5.1 The importance of data quality

It is important to ensure data quality for several reasons:

- solid data support conclusions and recommendations;
- future policy decisions may rely on the evidence generated in the survey;
- critics and opponents will look for weaknesses in the survey methods and results;
- results will be publicly accessible and may be used by others, e.g. in conducting international comparisons.

There are several reasons for data problems commonly encountered in a survey:

- field supervisors, data collectors and data entry personnel receive insufficient or poor-quality training;
- the pilot survey is not conducted properly;
- work in the field is of poor quality (insufficient supervision, no quality control for submission of completed forms, misunderstanding of instructions, etc.);
- data are not checked at every stage of the survey process;
- data are entered incorrectly;
- there are problems with uniquely identifying facilities;
- there are problems of human error;
- there is non-response to questions.

Data problems can therefore be avoided by:

- carefully studying the survey manual and accompanying materials at every step, and following instructions;
- selecting capable and reliable personnel and ensuring they are well trained in the survey methodology;
- encouraging personnel to communicate openly about uncertainties in survey procedures and questionable data;
- double-checking data collection forms for accuracy and completeness after each data collection visit, at the end of each day of fieldwork, and prior to data entry;
- conducting double entry of the survey data data are entered twice by different people and then crosschecked.

Thorough training of survey personnel is one of the most important ways of ensuring accurate data collection and good-quality data. Experience from previous surveys has shown that poor survey preparation, including inadequate training of survey personnel, results in onerous and time-consuming data checking that can significantly delay the survey's completion. It is therefore more effective and efficient to apply rigorous data collection methods than to try to clean or correct data once they are already collected.

1.5.2 Overview of training

All personnel involved in data collection, supervision and data entry require training to ensure reliable and accurate data collection, completion of questionnaires and data entry.

A training workshop for survey personnel should be held as part of the survey preparation. The overall objective of the training workshop is to provide field supervisors, data collectors and data entry personnel with the knowledge and skills required to carry out a SARA survey in an accurate and reliable manner.

Upon completion of the training, participants should:

- be familiar with the key aspects of the survey and how it is conducted;
- understand their roles and responsibilities in the survey, including specific tasks, timelines and reporting requirements;
- understand the critical content required to do their job effectively and possess the skills required to undertake each of their activities;
- be aware of common issues that may arise during survey activities, and understand troubleshooting/problem-solving strategies to address these issues;
- recognize the intrinsic value of good-quality data and be motivated to ensure data quality as part of their activities.

Training should therefore focus on teaching the following to the participants:

- the overall purpose of the survey;
- the consequences of poor-quality data;
- how to administer and record responses using the SARA questionnaire, the purpose and meaning of each question, and how to develop good rapport with the respondent;
- ethical issues involved in conducting a health facility survey, the importance of administering the informed consent statement, and how to maintain the privacy and confidentiality of the respondent;
- problem-solving in the field;
- how to enter data for both paper and electronic questionnaires;
- how to collect geographical coordinates of visited sites using GPS;
- common data collection and data entry mistakes.

It is recommended that the duration of a training workshop, which covers both data collection and data entry, last between 8 to 10 days (a sample agenda for the training of supervisors and data collectors is available in the *SARA implementation guide – Chapter 1: Planning*).

Training should include a data collection pilot test in which survey personnel visit public and private sector health facilities and collect data in the same way they would during actual fieldwork. This will not only provide survey personnel with practical experience in collecting data, but will also serve as a check of the appropriateness of the SARA questionnaire.

The trainer is usually the survey manager but could be a resource person with technical assistance from partner implementing agencies. The participants should include all field supervisors, data collectors and data entry personnel. For paper-based data collection, training on data entry is required. This can be held as a separate workshop or session for data entry personnel, however, there may be some advantage in holding a combined training session on data collection and data entry, since it will sensitize field supervisors and data collectors to the difficulties in entering poor-quality data. It is also recommended that the members of the survey coordinating group be invited to the introductory session of the training workshop to meet survey personnel and discuss the survey methodology.

The training workshop should be held as close as possible to the initiation of data collection – immediate departure for data collection can be scheduled if the survey manager has prepared well. Time lags between training and data collection should be avoided so that survey personnel have better recall of the data collection protocol.

1.5.3 Preparing for the training workshop

Planning the training workshop can require substantial time and preparation. Workshop preparations should begin early in the survey development process and should run in parallel to other survey planning and preparation activities. In preparing the training, it is essential to ensure that there is an adequate budget to cover costs for the training venue, transport, materials, and a daily allowance and accommodation for participants.

Select a training venue

A training venue should be selected based on the following criteria:

- availability of a room of appropriate size to hold the workshop;
- availability of essential technical resources (printer, photocopier, projector for presentations, electricity to charge mobile EDCs, etc.);
- proximity to health facilities that can be surveyed during the data collection pilot test;
- accessibility by routine modes of transport;
- on-site or nearby refreshments and accommodation for out-of-town participants;
- reasonable cost.

It is useful to check with survey coordinating group members to see if a meeting room can be made available for the training workshop at low or no cost.

Schedule dates of the training workshop

The training workshop should be scheduled close to the anticipated start of data collection. Do not plan the workshop during a time when weather or other conditions may delay the initiation of data collection. All survey personnel must attend the workshop and should be advised of the dates as early as possible. Invitations to attend the introductory session of the workshop should also be sent to survey coordinating group members.

Plan data collection pilot test

During the data collection pilot test, data collection team will visit at least three health facility (from different types) and collect data by following the survey procedures. It is recommended that each team visit one public health facility and one private health facility during the pilot test. The participation of pilot sites should be secured well in advance of the training workshops. The appointments should be made in advance and reconfirmed before the training session, avoiding peak periods when health facilities may be busy with patients.

Prior to the training workshop a written schedule should be prepared for each data collection team, indicating the time and location of each health facility visit, and including the name and contact details of the person in charge at the facility. The schedule should also contain the survey manager's telephone number so that survey personnel can call if there is a query or problem.

Secure equipment

All necessary equipment should be procured prior to the training session. This includes:

- projector, computer, etc. for the training session;
- pens, notepads, clipboards;
- mobile EDCs loaded with software and electronic forms;
- GPS devices for data collection teams;
- mobile phones for data collection teams to carry during the pilot test;
- access to a printer and photocopier for reproducing the SARA questionnaire.

Prepare training materials

Each training participant should receive:

- one copy of the SARA questionnaire;
- one copy of the SARA data collectors' guide;
- training hand-outs.

In addition, sufficient copies of the SARA questionnaire should be available for use in the pilot test.

1.5.4 Conducting the training workshop, including the data collection pilot test

The SARA data collectors' guide is available in the SARA Implementation Guide – Chapter 5: Data collector's guide. The guide provides:

- an overview of data collection processes;
- general guidance on interviewing practices and techniques;
- detailed explanations and definitions for each question in the questionnaire to provide a uniform understanding of the meaning of each question and response choices, and to improve the consistency of the data collected by different data collectors in different facilities.

This manual should be used during the training of all data collectors. In addition, slide presentations and accompanying hand-outs to complement *the SARA data* collectors' guide are available as tools for trainers to use during the training workshop.

The quality of data collection is controlled at several points in the data collection process. The first point of quality control is the thorough training of data collectors and the exclusion from fieldwork of any trainees who do not exhibit competency in applying the data collection questionnaires at the end of training.

Conducting the data collection pilot test

During the pilot test, data collection teams and their field supervisors will visit health facilities and collect data in the same way they would during the actual survey. Each field supervisor and data collector should complete their own SARA questionnaire to gain hands-on experience. Field supervisors should also supervise and watch out for common mistakes. It may be necessary to hold a preliminary pilot test with field supervisors to ensure that they are sufficiently knowledgeable about the survey protocol to supervise data collectors and identify mistakes. During the pilot test, any questions or uncertainties should be noted for clarification during the training workshop.

The data collection pilot test also serves as a pre-test of the questions in the SARA questionnaire and should help to highlight any country-specific adaptations that should be made to the survey including issues such as question format, wording and order. The pilot test allows for an opportunity to uncover any defects in the questions, glitches in wording of questions, lack of clarity of instructions, etc. The survey questionnaire should be piloted in all languages in which it will be administered. In addition to testing the paper questionnaire, the pilot test also tests field logistics, supervisory capacity and the application functionality for electronic data entry.

1.5.5 Finalizing the questionnaire

After piloting the SARA questionnaire, changes should be made to its format and/or content based upon any issues discovered during the piloting phase. All changes must be made to both paper and electronic versions of the questionnaire.

1.6 Preparing for data Collection in the field

The success of the SARA survey depends largely on the data collectors in the field, who are gathering and recording accurate, reliable data. Data collection requires careful planning and preparation, involving the following activities:

- planning the data collection visits
- preparing materials and tools for data collectors
- arranging transport and regular communications.

1.6.1 Planning the data collection visits

Who? Survey manager

The survey manager is responsible for planning the data collection visits. Before data collection starts, a schedule of visits to health facilities should be prepared for each survey area. The number of days required to collect the data can be estimated on the basis of the number of facilities to be visited in each geographical area, the distance between them and the mode of transport available. In general, two data collectors will require two hours plus travelling time for data collection in each facility.

Prepare a letter of introduction

Who? Survey manager

A letter of introduction from the survey manager is invaluable in introducing field supervisors – and later data collectors – to staff in the health facilities being surveyed. The survey manager should prepare a letter of introduction containing the following information:

- the name of the organization conducting the survey and the name of the survey manager
- contact details
- the purpose of the study
- the names of the data collectors who will visit the facility
- the time required for data collection in each facility.

The letter should also provide reassurance that the anonymity of the respondent will be maintained. The survey manager should provide field supervisors with sufficient signed copies for use during both the scheduling of field visits and the data collection visits.

Make initial contact with health facilities

Who? Field supervisors

It is essential that good relations be established with the person in charge of each facility to be surveyed, since they will have to set aside considerable time to provide information for the survey. Ideally, field supervisors should visit the heads of facilities personally, in advance, to seek their permission for data collection in their facility. Field supervisors should show them the letter of and introduction, and make an appointment for data collection on a date and at a time that is convenient for the head of the health facility, avoiding peak periods when he or she may be busy with patients. Field supervisors should note the contact person's name and telephone number at each health facility. If visits are not possible, then those in charge of the facility should be contacted by phone. The day before the scheduled data collection visit, field supervisors should telephone the health facility to confirm the appointment.

The following checklist should be used by field supervisors when contacting health facilities.

- Contact each health facility (sample and backup) to introduce the survey.
- □ Introduce the survey using the letter of introduction
- Make an appointment for data collection at a date and time that is convenient for the facility, avoiding peak hours. Allow two hours for data collection at a primary level facility, plus travel time. For larger facilities and hospitals, allow for additional time.
- □ Note the name and telephone number of the contact person at each health facility.
- Explain about the possibility of a second visit for 'validation', which should ideally take place in 10% of the sampled health facilities.
- **D** Before data collection starts, telephone each health facility to confirm the appointment.

Prepare a schedule of data collection visits

Who? Field supervisors

Field supervisors are responsible for preparing a written schedule for each data collection team. For each facility, the schedule should include the following:

- date and time of appointment
- name of facility
- contact person
- Iocation
- administrative unit
- unique ID number for the facility (provided by survey manager)
- name and contact details of a backup facility.

EXAMPLE OF A SCHEDULE FOR DATA COLLECTION VISITS

Survey area: Region 1

Data collection Team 1

Date/time of appointment	Name of facility	Contact person	Location	Managing authority	ID number	Backup site contact details
20 April 2012 10:00	ABC health centre	Mrs Nguyen	45 Main Street Eastern City Tel: +22 414 000	Private	01234	XYZ health centre 59 Main Street Eastern City Mr Shah

Data collection Team 2

Date/time of appointment	Name of facility	Contact person	Location	Managing authority	ID number	Backup site contact details

1.6.2 Preparing materials and tools for data collectors

Finalize and print questionnaire

Who? Survey manager

Following the data collection pilot test conducted as part of the training workshop, the survey manager should review and, if necessary, revise the SARA questionnaire. Both the paper and electronic versions of the questionnaire will need to be updated. Once the questionnaire has been finalized, the survey manager will need to print sufficient copies and also deploy the electronic forms to the mobile data collection devices.

Prepare data collection forms for each facility to be visited

Who? Field supervisor

The survey manager should provide the field supervisor with a separate questionnaire (data collection form) for:

- each sample health facility in the assigned survey area
- each backup facility
- each validation visit.

The survey manager should also provide the field supervisor with a list of the sample facilities in the survey area. Ideally, about 10% of the sampled facilities should be visited a second time for validation. The field supervisor will identify the validation sites by randomly selecting at least one public facility and one private facility from the list of sample facilities.

The field supervisor should prepare the data collection forms for each facility by completing the front page of the form (see Figure 1.6.1) with the identifying information of each sample facility, backup facility and validation facility, i.e. completing the following fields:

- name of health facility
- health facility unique ID
- name of town/village
- region and district
- type of facility
- managing authority

The following fields should <u>not</u> be completed by the field supervisor, as these will be completed by data collectors during the facility visits:

- date;
- name(s) of person(s) who provided information;
- name(s) of data collectors.

The verification at the top of the page should only be completed once the data collection form has been completed.

COVER	PAGE			
INTERVIE	EWER VISITS			
001	Is this a supervisor validation check of a facility?	YES	1 2	
Date Interview Name	1 ::	2 3 DA MC YEJ		
FACILITY	IDENTIFICATION			$\overline{)}$
002	Facility number			
	No me of to cility			
004	Location of facility			The field
005	Region/Province			supervisor should
005	District			complete this
007	Type of facility" "These should be adapted at country level prior to implementation"	NATIONA L REFERRA L HOS PITA L DISTRICT/PROVINCIA L HOS PITA L HEA LTH CENTRE/CUNIC HEA LTH POST MATERNA L/CHILD HEA LTH CUNIC OTHER S PECIFY]	1 2 3 4 5 96	distributing to data collectors
008	Meneging Authority (Ownership)	GOVERNMENT/PUBLC	1 2 3 4 96	
009	Urben / Rure I	URBAN	1 2	
010	Outpatient only	VES		

FIGURE 1.6.1: FRONT PAGE OF THE SARA DATA COLLECTION FORM

Arrange for storage of completed questionnaires

Who? Field supervisor and Survey manager

Field supervisors should arrange to store completed questionnaires until all fieldwork is completed, at which time they are transferred to the survey manager. A copy of all paper forms should be made by the field supervisor, and all paper forms should be stored in sealed plastic bags to prevent damage. Electronic forms should be synchronized daily to a central computer and a copy of all records should be stored on a memory card as backup. Field supervisors should always keep a copy of all data collection forms, in case those sent to the survey manager are lost or damaged. The survey manager should arrange for the safe storage of all completed forms in secure conditions for an indefinite period, in the event that data need to be checked at a later date.

Prepare materials and tools for data collectors

Who? Field supervisor

Data collectors need to bring tools and information with them on each day of data collection. Field supervisors should prepare resource kits containing all needed items for each data collection team. Before each day of data collection, the field supervisor should ensure that the data collectors have all the necessary tools and information with them including the following.

- □ A list of data collection teams and contact information.
- Contact details of the field supervisor, including a mobile phone number to call in case of difficulty in the field.
- □ A schedule of visits to survey sites.
- Contact details of the sites to be visited.
- Details of backup facilities to be visited if scheduled visits are not possible.
- **Copies of letter of introduction.**
- □ Data collector's guide and relevant hand-outs.
- □ A SARA data collection form for each health facility to be visited that day.
- □ A SARA data collection form for each backup site that may need to be visited that day.
- An EDC (fully charged and loaded with the SARA questionnaire), batteries and power cable
- □ A memory card for data backup (if applicable, depending on EDC selected) or USB key.
- □ A fully charged and accurately configured GPS unit.
- Pens (pencils should not be used to record data), a clipboard and other supplies.
- A notebook to record any significant events or findings.
- □ A field allowance for local expenses.
- □ An identity document with a photograph for each data collector.
- □ A mobile phone for each team and credit.

Where feasible, each data collection team should also be equipped with a mobile phone and credit to contact the field supervisor. Additional supplies may include a local map and extra batteries.

1.6.3 Arranging transport and regular communications

Arranging transport

Who? Survey manager or Field supervisor

Once all the survey sites are known, the survey manager or field supervisor should arrange transportation according to the number of sites to be visited, the number of teams going into the field, and the number of people per team.

Arranging regular communications

Who? Survey manager and Field supervisor

Throughout the data collection process, field supervisors should be available to provide advice to data collectors and answer any questions they may have. Providing data collectors with their field supervisor's mobile phone number, when feasible, is one way of ensuring timely communication.

Data collectors should also meet with their field supervisor on a regular basis so that completed forms can be checked and any issues can be resolved. Ideally, this should occur at the end of each day of data collection so that errors do not carry over into future data collection visits. In addition, data collectors will be better able to recall the data collection visit, which may be useful in clarifying erroneous or illegible data. During data collection, data collectors should record how problems were solved or how data collection was simplified. These notes should be reviewed with the field supervisor during the debriefing.

The survey manager should also be available throughout the data collection process to respond to questions from field supervisors, and the survey manager should provide field supervisors with his/her mobile phone number for this purpose. Ideally, the survey manager should visit each survey area during data collection to supervise activities. If this is not possible, he or she should arrange for regular communications with each field supervisor to receive updates on the data collection process.

1.7 Data collection in the field

This chapter describes procedures for data collection in the field. Table 1.7.1 shows the activities involved for each day of data collection.

TABLE 1.7.1:	DAILY	ACTIVITIES	FOR	DATA	COLLECTI	ON
--------------	-------	------------	-----	------	----------	----

When?	What?	Who?	
Before going out to collect data each day	Check that the data collection teams have all the materials necessary for field visits and confirm transport arrangements	Field supervisors/data collectors	
	Call each facility to be visited and confirm appointment	Field supervisors	
On arrival at the facility	Introduce survey team and remind facility staff of the purpose of the visit	Data collectors	
	Verify and complete the SARA questionnaire	Data collectors	
	Check that all data are entered on the SARA questionnaire before leaving the facility.	Data collectors	
At the end of each day	Conduct meeting between field supervisors and their data collectors, and discuss any difficulties	Field supervisors/data collectors	
	Review each SARA questionnaire and clarify missing/unreliable information	Field supervisors	
	Sign, copy and store all checked data collection forms	Field supervisors	

Each step of data collection is described below according to the personnel responsible, namely field supervisors and data collectors.

1.7.1 Field supervisors: fieldwork responsibilities

Field supervisors are responsible for ensuring the accuracy and reliability of data collection. This involves the following activities.

Field supervision

Field supervisors should meet with their data collectors at the end of each day to check completed data collection forms, get feedback on the data collection process and resolve any problems. They should visit the health facilities regularly with the data collection teams to ensure that the agreed procedures are being followed.

Daily check of completed SARA questionnaires

It is important that field supervisors review completed SARA questionnaires at the end of each day to check that the data are complete, consistent and legible. Once the team has left the field, it becomes difficult to verify information that may be missing or incomplete.

The supervisors should highlight any missing or unreliable information on the form and identify the source of the problem. If necessary, data collectors should return to the facility to collect any further data required. Once the field supervisor is satisfied with the completeness and reliability of a SARA questionnaire, he or she should

sign the form in the designated place to record that it has been checked. Forms should then be safely stored until completion of data collection, at which time they are transferred to the survey manager.

Validation of data collection

Field supervisors should validate data collection by repeating the survey, using a "light" version of the questionnaire (sub-selection of sections) at selected health facilities (about 10%) and checking their results against those of their data collectors. Where possible, health facilities visited for validation should be selected at random. Ideally, the validation should be done on the same day as data collection (soon after the data collectors have left the facility) to avoid changes in the availability of the items. Any discrepancies between the results of the field supervisor and those of their data collectors should be discussed with the data collectors, and the data collection protocol should be clarified as necessary. Any problems that cannot be resolved in the field should be discussed with the survey manager.

This validation can also be conducted by the entity in charge of the survey quality assurance. The supervisors ease access to the data collected by the data collection teams in view of the data files comparisons.

Storing of completed SARA questionnaires

Completed paper questionnaires should be copied and stored in sealed waterproof plastic bags, in a location that is protected from moisture, direct sunlight, rodents and insects. Originals should be stored in a separate location from copies. Electronic questionnaires should be synchronized with a central computer and saved both on the computer hard drive and on an external memory card for safe keeping. All original data collection questionnaires, including those for validation visits, should be transferred to the survey manager upon completion of fieldwork. Field supervisors should retain the copies for use in the event that the originals become lost or damaged.

In order to accomplish these tasks, each field supervisor should have the following materials:

- □ A full list of sample sites (and backup sites) for survey area and contact details.
- □ An assignment of sites by data collection team.
- □ A list of data collection teams and contact information when in the field.
- □ A schedule of visits to survey sites and contact details of the sites.
- **Copies of letter of introduction.**
- **C** Copies of the supervisor and data collector's guides and other relevant documents/material.
- **Extra copies of the SARA data collection form.**
- A data collection form for data validation at each facility that may need to be visited that day.
- □ A fully-charged laptop computer with appropriate software (CSPro)
- Extra EDCs as backup (fully charged and loaded with the SARA questionnaire) in case of loss or damage, with extra batteries and power cables.
- Extra memory cards for data backup or USB keys (depending on the EDC used).
- Extra GPS units as backup (fully charged and accurately configured).
- Pens (pencils should not be used to record data), a clipboard and other supplies.
- □ A notebook to record any significant events or findings.
- □ A field allowance for local expenses.
- □ An identity document with a photograph.
- A cell phone with credit

1.7.2 Data collectors: fieldwork responsibilities

Before visiting the facilities each day

Before visiting the facilities each day, data collectors should check that they have all the materials they will need for data collection.

- □ A list of data collection teams and contact information.
- Contact details of the field supervisor, including a mobile phone number to call in case of difficulty in the field.
- □ A schedule of visits to survey sites.
- □ Contact details of the sites to be visited.
- Details of backup facilities to be visited if scheduled visits are not possible.
- **Copies of letter of introduction.**
- □ Data collector's guide and relevant hand-outs.
- □ A SARA data collection form for each health facility to be visited that day.
- □ A SARA data collection form for each backup site that may need to be visited that day.
- □ An EDC (fully charged and loaded with the SARA questionnaire), batteries and power cable
- □ A memory card for data backup (if applicable, depending on EDC selected) or USB key.
- □ A fully charged and accurately configured GPS unit.
- Pens (pencils should not be used to record data), a clipboard and other supplies.
- □ A notebook to record any significant events or findings.
- □ A field allowance for local expenses.
- □ An identity document with a photograph for each data collector.
- □ A mobile phone for each team and credit.

Where feasible, each data collection team should also be equipped with a mobile phone and credit to contact the field supervisor. Additional supplies may include a local map and extra batteries.

On arrival at the facility

On arrival at the health facility, data collectors should do the following.

- Introduce themselves and remind health facility staff of the survey's purpose as well as the scheduled data collection visit. Data collectors should also thank the staff for their cooperation and, if necessary, remind them that the respondents' identity will be kept confidential.
- Check that the facility information on the first page of the SARA questionnaire is complete and correct, informing the field supervisor at the end of the day if there were any inaccuracies.
- Fill in the date and names of the data collectors on the cover page.
- Take the GPS coordinates of the health facility.
- Obtain informed consent to begin the survey.
- Fill out the SARA questionnaire making sure to speak to the most knowledgeable person in the health facility for each section of the questionnaire. One data collector should complete the SARA data collection paper form and another should complete the SARA electronic form, paying close attention to the instructions on the forms. Data collectors should not leave the SARA data collection form at the facility to be filled in later. A separate SARA data collection form should be completed at each facility.

Before leaving the facility

Before leaving the health facility, data collectors should do the following.

• Double-check that the data collection form is legible, accurate and complete.

NOTE: Backup facilities to be visited are identified in the schedule. The field supervisor should determine when it is necessary to visit a backup facility. When visiting a backup facility, questionnaires should be completed in exactly the same way as in other facilities, making sure to complete the SARA data collection form that corresponds to that facility.

• Thank staff at the facility for their participation.

At the daily meeting with the field supervisor

At the end of each day, data collectors should meet with the field supervisor and do the following:

- 1. back-up electronic data on the memory card in the EDC
- 2. submit the data collection forms and files completed that day
- 3. transfer data from the EDC to the central computer
- 4. report on the activities of the day
- 5. recharge the battery of the EDC to be ready for the next day
- 6. check the battery life of the GPS unit and get a second set of batteries if necessary
- 7. recharge mobile phone if necessary.

Data collectors should alert their field supervisor of any problems or uncertainties regarding data collection procedures. They should also report any problems with electronic equipment and arrange to get replacements if necessary.

1.7.3 Ensuring data quality

The quality of the information that the SARA survey generates depends on the accuracy of data collection. The survey manager has overall responsibility for the quality of the data, although all survey personnel have a role to play in ensuring the accuracy of the data collected. The field supervisors and data collectors should receive regular supervision. Rigorous enforcement of data collection procedures will pay off with the ease with which data entry and analysis occur. The following steps will also help to ensure greater accuracy of data collection.

- 1. Ensure that there is thorough preparation and training as a first step in minimizing errors.
- 2. Establish procedures to check for data completeness, consistency, plausibility and legibility in the field when it is still possible to correct errors or to fill in missing information. Field supervisors should review data collection forms every day after completion of the fieldwork and resolve any problems before the next day of data collection.
- 3. Plan random checks to ensure the quality of data collection. It is recommended that an entity returns to randomly-selected health facilities (10%) to collect the same data so as to check the accuracy of the first data set. Ideally, the validation should be done on the same day as data collection (soon after the data collectors have left the facility) to avoid changes in the availability of survey items.
- 4. Double-check all completed SARA questionnaires; verify any suspicious, incomplete or illegible data prior to the initiation of data entry.

1.8 Data entry and processing

If data are collected on paper forms, they must be entered electronically before proceeding with data processing and analysis. If data are collected in electronic forms, then one can proceed directly to the data processing step.

Once in electronic form, the data need to be checked for accuracy, completeness and consistency before the data set can be finalized. Any errors or inconsistencies must be flagged and resolved prior to analysis. The purpose of editing is to eliminate omissions and invalid entries, e.g. by changing inconsistent entries, and should be kept to a minimum: data should never be changed to conform to expectations. It is good practice to always preserve an unedited copy of the data set and to document in detail the data editing process.

Finally, once the data have been checked and verified, it is customary to export the final data set in some commonly-used file format, such as a spreadsheet file format or CSV (comma-separated values). This is useful for sharing the data with other parties, and to perform analysis in other statistical software packages.

1.8.1 Data entry

Any data collected on paper must be entered electronically before it can be processed and analysed. With input from technical members of the survey coordinating group, the survey manager selects the appropriate data entry software and sets up a data entry operation. Transferring the data from paper to electronic form can be a source of error; therefore, it is important to have the appropriate data validation processes in place to ensure accurate data entry. If electronic data collection has been used, the data already exist in an electronic format and this step is complete.

Selecting data entry software

When selecting data entry software, there are two main principles to consider:

- 1. use software that speeds up data entry and minimizes errors
- 2. have a thorough knowledge of the software selected.

Keeping these principles in mind when thinking about data entry software options helps to narrow down the potential options and results in selection of an appropriate solution.

While it is possible to use many types of software for data entry (including statistical programs, database management systems and spreadsheets), it is recommended that a specialized data entry software such as CSPro be used to minimize the possibility of entry errors and to facilitate validation.

Statistical software

Statistical software package programs are software packages that are specialized for data analysis. Some include data entry and data checking functions in addition to data analysis (e.g. CSPro), while others are useful primarily for data analysis and visualization. Some advantages of using a software package with built-in data entry and verification functionality are (1) data entry clerks are less likely to make mistakes when entering data, and (2) mistakes are much easier to identify and fix. In particular, the software can be programmed to provide a highly-structured data entry environment so that only valid values are accepted and skip patterns⁵ are automatically integrated. In addition, such software facilitates independent data verification, in which the data are entered manually twice and differences are later reconciled. Once the data have been entered, it may be

⁵ Skip patterns are a particular type of survey branching logic that will jump a respondent over a group of questions that isn't relevant to them.

necessary to use a different statistical software package to perform analysis, depending on whether the package offers the desired data analysis and graphing functions. These types of tools require some advanced technical knowledge, but overall result in improved data quality.

For SARA, the recommended data entry software is CSPro: a statistical software package with built-in data collection/entry functionality that allows for speedy data entry while also providing sufficient checks and data validation to ensure quality data. CSPro includes all the necessary functionality for SARA and can be downloaded free of cost.

Database management systems

A database management system is an application that allows the creation and management of databases, including storage and retrieval of data. There are different types of databases, but the most popular is a relational database that stores data in tables where each row ("record") in the table holds the same sort of information. Each record has a unique identifier ("key"), which allows retrieval of information from related tables. Databases are more difficult to set up than spreadsheets, but they allow more sophisticated data retrieval and search. In addition, it is possible to write scripts using a query language such as SQL to perform simple data checking and other functions. However, these programs are designed mainly for data storage and retrieval, and are usually not designed to facilitate manual data entry. For the SARA survey, it is recommended that a database system be used to store the data once it has been entered, but not for data entry itself.

Spreadsheets

Spreadsheet programs offer the most basic option for data entry. Although spreadsheets are easy to use and many people are knowledgeable about how to use the software programs, there are many disadvantages of using spreadsheets for data entry: it is very easy to make a mistake, data entry is slow, and there is no built-in checking for valid values. As a result, for the SARA survey, it is not recommended to use a spreadsheet for data entry unless there are no other viable options. Spreadsheet software is often useful to view data once it has been digitized and stored in a database.

Preparing for data entry

The data entry application must be designed using the selected data entry software. Valid values should be defined for certain responses and entry of data should be restricted to these values alone. Furthermore, special keys for missing data should be included in the value set and may use a standard identifying digit. Open-ended questions or the selection of the broad category of "other" can also be programmed to allow for entering the written response. This keying of open-ended questions will require the manual coding of these responses at some future date.

A centralized system for data entry should be set up, with one or more groups of data entry clerks managed by a supervisor. The number of data entry clerks will depend on a number of factors, including (1) budget, (2) timeline, (3) the availability of qualified personnel, and (4) the availability of computers and other equipment for data entry. Generally, the more data entry clerks there are, the quicker the data can be entered.

Part of the management and organization of a data entry operation requires establishing a specified work schedule. Monitoring the productivity of the individual data entry clerks should be part of a data entry system as well. Like other process, the data entry process requires good organizational and project management skills.

Entering data

Data should be entered using the software that has been selected.

In the data entry process, it is important to consider the following issues.

Missing data

In general, it is not good practice to use blanks as missing data codes. Missing data can arise in a number of ways, and it is important to distinguish among these different instances. There are at least five missing data situations, each of which should have a distinct missing data code.

- **Refusal/no answer**. The subject explicitly refused to answer the question or did not answer the question when he or she should have.
- Don't know. The subject was unable to answer the question, either because he or she had no opinion or because the required information was not available (e.g. a respondent could not provide information on the functionality of equipment due to inaccessibility).
- **Processing error**. For some reason, there is no answer to the question although the subject provided one. This can result from interviewer error, incorrect coding, machine failure or other problems.
- Not applicable. For one reason or another, the subject was never asked the question. Sometimes this results from "skip patterns"⁶ that occur (e.g. for facilities that do not have a generator, questions regarding generator functionality and availability of fuel would be not applicable).
- **No match**. This situation may arise when data are drawn from different sources (e.g. a survey questionnaire and an administrative database), and information from one source cannot be located.

Selecting missing data codes

Missing data codes should always match the content of the field. If the field is numeric, the codes should be numeric, and if the field is alphabetic, the codes may be numeric or alphabetic. Most researchers use codes for missing data that are above the maximum valid value for the variable (e.g. 97, 98 and 99). Missing data codes should be standardized so that only one code is used for each missing data type across all variables in the data file or across the entire collection if the study produced multiple data files.

Not applicable and skip patterns

Handling skip patterns is a constant source of error in both data management and analysis. On the management side, deciding what to do about codes for respondents who are not asked certain questions is crucial. "Not applicable" codes, as noted above, should be distinct from other missing data codes. It is not good practice to leave the record blank. Data set documentation should clearly show for every item exactly who was asked and who was not asked the question. At the data cleaning stage, all "filter items" should be checked against items that follow to make sure that no one provides answers to the item who should not, and that those who did not answer the item have the correct kind of missing data code.

⁶ Skip patterns are a particular type of survey branching logic that will jump a respondent over a group of questions that isn't relevant to them.

1.8.2 Data processing

After data entry, data should be checked for inconsistencies and possible errors. If the data are collected electronically, field supervisors should, for the duration of the data collection phase, check data for all health facilities that were visited each day. If the data are transferred from paper to electronic versions, the data should be checked for entry errors. It is particularly important to check that the facility ID items such as the facility number, name, location, facility type and managing authority have been entered correctly, and that there are no inconsistent or missing data. Usually errors can be resolved by reviewing all of the information provided by a respondent or by referring to the paper copy of the questionnaire responses.

Edit and correct data

The purpose of editing is to make the data as representative of the real life situation as possible; this can be done by eliminating omissions and invalid entries, and by changing inconsistent entries. Below are some important principles that should be followed.

- The fewest number of changes should be made to the originally recorded data. The goal is to make a record or questionnaire acceptable, not to make it conform to what one thinks should be acceptable.
- For certain items it may be acceptable to have a "not reported [NR]" or "not stated [NS]" category. Thus, in case of an omission or an inconsistent, impossible or unreasonable entry, a code for "NR" or "NS" can be assigned.
- Obvious inconsistencies among the entries should be eliminated.
- Providing corrected values for erroneous or missing items should be supplied by using other values as a guide, and always in accordance with specified procedures.
- Specifications for editing the questionnaire data should be developed at the same time as the questionnaire itself.

Remove any duplicate records

It is possible that a facility has been entered in the database twice and thus the duplicate record must be removed. For any records that are identical, one should be removed. If two records appear to be duplicates according to facility name, but do not contain the same data, a list of criteria must be used to determine if it is a true duplicate. The following data elements could be used as the criteria for determining duplicates:

- district
- facility code/name
- GPS coordinates
- facility type
- managing authority
- interviewer's name.

If these are all the same it is safe to consider the records as duplicates. At this point, the most complete record should stay in the data set. If both records are complete, the record with latest time stamp should be kept.

Check validity of GPS coordinates

GPS coordinates should be checked to ensure that they fall within the boundaries for the country and region. Sometimes latitude and longitude coordinates can be entered incorrectly (they can be inversed and +/- signs can be reversed, or an incorrect format can be entered). All GPS coordinates should be double-checked to ensure they are valid for the area being surveyed.

Check validity of responses

Data entry software often has built-in functionality to check data as it is being entered, such as range checks and within-record consistency checks. Data editing programs can be written to check the validity of responses after data entry, including whether the data follow the appropriate skip patterns.

Recode values for "other"

Questions where "other" is a possible response option should be checked, and the written responses reviewed to determine if the response actually corresponds with one of the pre-coded options. If this is the case, these responses should be recoded to the appropriate response category.

Review comments sections

At the end of the survey, there are several questions allowing for the interviewer to provide comments. Please review these sections for any relevant information.

Data validation and verification

Verification is a process of double entry of the same questionnaire and comparing the responses. This can either be the paper questionnaire entered twice or validation between paper and electronic versions if electronic data collection is used with paper questionnaires as a backup. Differences in keyed data of the same questionnaire need to be reconciled. A system of verification can virtually assure that the information presented in the questionnaire is faithfully keyed. Verification can be dependent or independent. Dependent data entry uses one data file and reconciles any identified error with the original data file. Independent verification is the process of keying to fully independent data files of the same questionnaire or cluster and comparing the two files. A report of inconsistencies is issued and the differences between the two data files must be fully reconciled.

Data clean-up

Before finalizing and exporting the data set, the following steps should be taken to clean the data set as applicable.

Rename the variables

The variable should be named according to the corresponding question number in the survey. This may already be the case if the database is set up in this way. If electronic software is used, variables are often assigned names based on category headings, which are sometimes long and cumbersome to use and do not provide a good description of the variable and thus require renaming. For example, the variable "_2_001_date" which corresponds to question 001 in the survey will be renamed "q001".

Label the variables

Adding a label to a variable allows a text description to be associated with the variable name. For example, the new variable "q001" can have a label called "date." This enables the user to more easily identify what each variable represents.

Remove variables for which no data exist

If data are collected using electronic software, there may be variables in the data set which are actually instructions from the questionnaire and do not include any data. These variables must be removed from the data set.

Define the data type associated with variables

There are generally two data types associated with variables: numeric and string. Numeric variables are simple: they contain numbers. String variables contain text that can contain any characters on the keyboard: letters, numbers and special characters. It is important to define each variable according to the appropriate type in order for statistical analysis to be carried out on the data.

Adjust variables in which two numeric responses have been chosen

All variables with numeric responses should contain only one response. It is possible that a single select numeric variable is erroneously assigned two values. In some programs, this is represented in the data set as #;# and causes the variable to be categorized as a string variable due to the non-numeric character. These values must be imputed so that only one numeric value is recorded and then the data type of the variable must be converted from a string to a numeric value.

1.8.3 Exporting the data set

Once the data set has been processed and verified, it is good practice to export the finalized data set into some commonly used file format such as a spreadsheet format or CSV (.csv). This is useful when sharing the data and for analysis of the data using statistical software packages.

1.9 Data analysis

Once data have been verified, data analysis can begin. There are many different types of results that can be obtained from surveys. The types of analysis used depend to a large extent on the design determined in the planning phase of the SARA survey. Some data analyses are standard and are included in most survey reports. However, not all of the analyses of the survey data need to be included in the final report, as the focus should be on the most important and relevant results. Therefore, survey managers should generate the full range of survey results, and together with the survey coordinating group, select the most significant findings for inclusion in the final report. It is only by conducting a complete analysis of the survey data that it can be assured that important findings have not been overlooked. Based on the initial set of results from the standard analyses, there is often further analysis in areas of interest. Following data analysis, a meeting with the survey coordinating group should be held to assist in interpreting the results and developing recommendations.

Survey indicators are important in providing crucial information for informed policy choices, especially to decision-makers, programme planners and policy-makers. Serving as baselines, indicators are important for setting goals and targets for the future and allow for a certain level of comparability between surveys of different location and time period. Moreover, indicators help place focus on predetermined areas of a survey that are deemed to be most useful, relevant and important to the current health system. Having a consistent indicator set also contributes to standardized analytical reporting.

SARA uses both tracer indicators and composite indicators in data analysis. Tracer indicators aim to provide objective information about whether or not a facility meets the required conditions to support provision of basic or specific services with a consistent level of quality and quantity. Summary or composite indicators, also called indices, are a useful means to summarize and communicate information about multiple indicators and domains of indicators. Composite indices are useful to help get an overall view of the situation and to summarize multiple pieces of information. For SARA, composite indices are useful to compare districts or regions or to look at change over time. However, composite indices also have limitations. It can be difficult to understand the individual factors contributing to an index score, and thus it is important to have information on individual indicator items in addition to composite index scores.

The following sections provide an overview of how to calculate SARA indicators and indices.

1.9.1 Calculating the service availability indicators and index

Overview

An important note regarding service availability: although this information is collected through the SARA questionnaire, these indicators should not be calculated for a sample of facilities. **Data must be available for ALL facilities in an administrative unit in order to calculate service availability.** All service availability measures require data that link the numerator (e.g. number of facilities) to the denominator (e.g. population size). A sample survey would not allow computation of the service availability indicators as it is not clear what the corresponding population size to be used as the denominator should be.

The information needed to calculate service availability can be gathered from multiple sources in addition to the SARA questionnaire, namely the HMIS and other routine information systems, and should be collated for all facilities before calculating the service availability indicators. If SARA is implemented as a census, then it can be used to calculate service availability.

Service availability is described by three domains of tracer indicators: health infrastructure, health workforce and service utilization.

Health infrastructure indicators

- Facility density (number per 10 000 population): the facility density is primarily an indicator of outpatient service access.
- Inpatient bed density (number per 10 000 population): inpatient bed density provides an indicator of the inpatient services access. Paediatric beds (cots) are included, but maternity beds are excluded.
- **Maternity bed density (number per 1000 pregnant women):** maternity bed density provides an indicator of access to delivery services. Data on maternity beds can be used calculate the density of maternal beds per 1000 pregnant women per year. The denominator is estimated from the population data. The indicator does not include delivery beds.

Health workforce indicator

• Health workforce density (number per 10 000 population): the health workforce density is the number of core medical professionals per 10 000 population: physicians, non-physician clinicians, registered nurses and midwives. This includes part-time physicians who are given the value of 0.5 in the scoring.

Service utilization indicators

In populations with poor or suboptimal health infrastructure, the service utilization rate is an indicator of access.

- **Outpatient service utilization (number of outpatient visits per capita per year):** the number of visits for ambulant care, not including immunization, over the total population.
- Inpatient service utilization (number of hospital discharges per 100 population per year, excluding deliveries): this indicator provides additional information on the availability and access to inpatient services.

These indicators must all be expressed as a percentage score compared with a target or benchmark. Table 1.9.1 shows the target and computation of each indicator. If the tracer indicator score exceeds the target, it is scored as 100%.

Domain		Indicator	Target*	Score (%) (n / target, maximum 100)		
Health infrastructure						
а	Facility density	Number per 10 000 population (n)	2	n / 2 × 100		
b	Inpatient bed density	Number per 10 000 population (n)	25	n / 25 × 100		
с	Maternity bed density	Number per 1000 pregnant women (n)	10	n / 10 × 100		
Не	Health workforce					
d	Core health workforce density	Number per 10 000 population (n)	23	n / 23 × 100		
Se	Service utilization					
е	Outpatient service utilization	Outpatient visits per person per year (n)	5	n / 5 × 100		
f	Inpatient service utilization	Hospital discharges per 100 per year (n)	10	n / 10 × 100		

Health infrastructure targets and scores

The rationale for the targets can be summarized as follows.

Facility density (*a*): usually there is a country target, such as at least one facility per 5000 population, or two facilities per 10 000 population. A major limitation is that this indicator does not take into account the size of the facilities. The indicator is scored as $n / 2 \times 100\%$ (maximum 100), where n is the number of facilities per 10 000 population.

Inpatient bed density (*b*): the global average is 27 per 10 000 (*10*). Lower- and upper-middle-income countries have 18 and 39 hospital beds per 10 000, respectively (*10*). For SARA, an arbitrary benchmark of 25 per 10 000 is selected. The indicator is scored as $n / 25 \times 100\%$ (maximum 100), where n is the number of inpatient beds per 10 000 population.

Maternity bed density (c): under the assumption that there should be sufficient beds for all pregnant women with an occupancy rate of 80% (to account for the uneven spread of demand over time) and a mean duration of stay of 3 days, the target should be $(1000 / 0.8) \times (3 / 365) = 10$ per 1000 pregnant women. The indicator is scored as $n / 10 \times 100\%$ (maximum 100), where n is the number of maternity beds per 1000 pregnant women.

An estimation for the number of pregnant women in the population can be derived from the CBR (crude birth rate) for the country of interest and the following equations:⁷

- A = estimated number of live births = (CBR per 1000 × total population)
- B = estimated live births expected per month = (A / 12)
- C = estimated number of pregnancies ending in stillbirths or miscarriages = $(A \times 0.15)$
- D = estimated pregnancies expected in the year = (A + C)
- *E* = estimated number of women pregnant in a given month = $(0.70 \times D)$
- F = estimated % of total population who are pregnant at a given period = (E / total population × 100).

Health workforce target and score

Health worker density (*d***):** The published figure by WHO is 23 per 10 000 population (9). The indicator is scored as $n / 23 \times 100\%$ (maximum 100), where n is the number of core health workers per 10 000 population.

Service utilization targets and scores

Outpatient service utilization (*e*): in countries of the Organisation for Economic Co-operation and Development (OECD), the average number of physician consultations per person per year is about six (10). For SARA, the proposed benchmark is five visits per person per year. The indicator is scored as $n / 5 \times 100\%$ (maximum 100), where n is the number of outpatient visits per person per year.

Inpatient service utilization (*f*): in OECD countries, which have an ageing population, there are about 15 discharges per 100 population per year (*11*). For SARA, the proposed benchmark is 10 discharges per 100 people per year. The indicator is scored as $n / 10 \times 100\%$ (maximum 100), where n is the number of hospital discharges per 100 people per year.

The service availability index is calculated using the six above mentioned indicators. First, indices are calculated for health services infrastructure, health workforce and service utilization. The calculations for creating those indices are shown in Table 1.9.2 (please refer Table 1.9.1 for the definitions of indicators a-f). The service availability index is the unweighted average of the three areas: infrastructure, health workforce and utilization: [((a + b + c) / 3) + d + ((e + f) / 2)] / 3, and is a percentage score.

⁷ These equations can be found at:

http://www.who.int/reproductivehealth/publications/emergencies/field manual rh humanitarian settings.pdf, Chapter 5, Annex 3.

TABLE 1.9.2: SERVICE AVAILABILITY INDICES

Index	Indicator Target		Score	
Health infrastructure index	Average score of the three indicators: facility density, inpatient bed density, maternity bed density	100	(a + b + c) / 3	
Health workforce index	Core health worker density	100	d	
Service utilization index	Average score of the two indicators: outpatient visits, hospital discharges	100	(<i>e</i> + <i>f</i>) / 2	
Service availability index	Unweighted average of the three areas: infrastructure, workforce and utilization	100	[((a + b + c) /3) + d + ((e + f) / 2)] / 3	

Required data sources

Table 1.9.3 shows the required information and potential data sources for calculating service availability.

TABLE 1.9.3: DATA SOURCES

Information needed	Potential data source
List of all health facilities	MFL
Service utilization data	HMIS
Health workforce data	Human resources information system (HRIS)
Inpatient and maternity beds data	Varies by country
Population data (national and regional/district depending on how results will be reported)	National Bureau of Statistics

Example calculation

Table 1.9.4 shows the data used for this example.

TABLE 1.9.4: EXAMPLE DATA

Data item	Value
Number of facilities	400
Number of inpatient beds	5500
Number of maternity beds	800
Number of core health workers	4600
Number of outpatient visits per year	9 000 000
Number of hospital discharges per	225 000
Population	3 000 000
Crude birth rate (CBR)	40

There are three main steps to calculate the service availability index.

Step 1. Calculate service availability indicators

The first step is to calculate the six service availability indicators. The following example (Table 1.9.5) shows the equations used to calculate each of the six indicators using the example data values.

TABLE 1.9.5:	CALCULATING THE INDICATORS
--------------	-----------------------------------

Indicator	Value
Facility density (number per 10 000 population)	number of facilities / population = <i>n</i> / 10 000 400 / 3 000 000 = <i>n</i> / 10 000 <i>n</i> = 1.33
Inpatient bed density (number per 10 000 population)	number of inpatient beds / population = <i>n</i> / 10 000 5500 / 3 000 000 = <i>n</i> / 10 000 <i>n</i> = 18.33
Maternity bed density (number per 1000 pregnant women)	number of maternity beds / pregnant population* = $n / 1000$ 800 / 96 600 = $n / 1000$ n = 8.28 *see Table 1.9.6 for how to calculate number of pregnant women
Health workforce density (number per 10 000 population)	number of core health workers / population = <i>n</i> / 10 000 4600 / 3 000 000 = <i>n</i> / 10 000 <i>n</i> = 15.33
Outpatient service utilization (outpatient visits per capita per year)	number of outpatient visits per year / population = <i>n</i> 9 000 000 / 3 000 000 = <i>n</i> <i>n</i> = 3.00
Inpatient service utilization (hospital discharges per 100 population, excluding deliveries)	number of hospital discharges per year / population = $n / 100$ 225 000 / 3 000 000 = $n / 100$ n = 7.50

(40 / 1000) x 3 000 000 = 120 000
120 000 / 12 = 10 000
120 000 x 0.15 = 18 000
120 000 + 18 000 = 138 000
0.7 x 138 000 = 96 600
(96 600 / 3 000 000) x 100 = 3.22

Service availability indicators can each be displayed in a graph such as the one for health workforce density in Figure 1.9.1.



FIGURE 1.9.1: CORE HEALTH WORKERS PER 10 000 POPULATION

Step 2. Calculate service availability indicator scores

Next, use the values obtained from Step one to calculate the service availability indicator scores. The scores compare the indicator to a target and are expressed as a percentage. Table 1.9.7shows the calculations for each of the six service availability indicator scores.

TABLE 1.9.7: CALCULATING THE SERVICE AV/	AILABILITY INDICATOR SCORES
--	-----------------------------

Domain		n	Target	Score (%)		
				(n / target) x 100 (maximum 100)		
Health infrastructure						
а	Facility density	1.33	2	(1.33 / 2) x 100	66.5	
b	Inpatient bed density	18.33	25	(18.33 / 25) x 100	73.3	
с	Maternity bed density	8.28	10	(8.28 / 10) x 100	82.8	
Health workforce						
d	Core health workforce density	15.33	23	(15.33 / 23) x 100	66.7	
Service utilization						
е	Outpatient service utilization	3.00	5	(3 / 5) x 100	60.0	
f	Inpatient service utilization	7.50	10	(7.5 / 10) x 100	75.0	

Step 3. Calculate service availability indices

Lastly, use the service availability indicator scores to create the health infrastructure index, the health workforce index, the service utilization index and the overall service availability index. Table 1.9.8 shows these four index calculations using the example data.

TABLE 1.9.8:	CALCULATING THE SERVICE AVAILABILITY INDEX
--------------	---

Index	Indicator	Score (%)	
Health infrastructure index	Average score of the three indicators: facility density, inpatient bed density, maternity bed density	(<i>a</i> + <i>b</i> + <i>c</i>) / 3	(66.5 + 73.3 + 82.8) / 3 = 74.2
Health workforce index	Core health worker density	d	66.7
Service utilization index	Average score of the two indicators: outpatient visits, hospital discharges	(e + f) / 2	(60.0 + 75.0) / 2 = 67.5
Service availability index	Unweighted average of the three areas: infrastructure, workforce and utilization	[((a + b + c)/3) + d + ((e + f) / 2)] / 3	(74.2 + 66.7 + 67.5) / 3 = 69.5

The service availability indices can be displayed in a graph such as the one in Figure 1.9.2.



FIGURE 1.9.2: SERVICE AVAILABILITY INDICES

1.9.2 Calculating the general service readiness indicators and index

Overview

General service readiness is described by the following five domains of tracer indicators:

- Basic amenities
- Basic equipment
- Standard precautions for infection prevention
- Diagnostic capacity
- Essential medicines.

Each domain consists of a set of tracer items. Table 1.9.9 lists the tracer indicators for each domain.

General service domains	Tracer items	Domain score	
(a) Basic amenities	• Power	$n / 7 \times 100$, where <i>n</i> is the total	
(a) Basic amenices	Improved water source facility promises	number of items available in the	
	Recommute auditory and visual privacy for nations		
	consultations		
	Access to adequate sanitation facilities for clients	-	
	Communication equipment (phone or short-wave radio)	-	
	Access to computer with e-mail and Internet	_	
	Emergency transportation		
(b) Basic equipment	Adult scale	$n / 6 \times 100$ where <i>n</i> is the total	
	Child scale	domain	
	• Thermometer		
	Stethoscope		
	Blood pressure apparatus		
	Light source		
(c) Standard	Safe final disposal of sharps	$n / 9 \times 100$ where <i>n</i> is the total	
precautions for infection prevention	Safe final disposal of infectious wastes	domain	
	 Appropriate storage of sharps waste (sharps box/container) 		
	 Appropriate storage of infectious waste (waste receptacle with lid and plastic bin liner) 		
	Disinfectant		
	• Single-use, standard disposable or auto-disable syringes		
	Soap and running water or alcohol-based hand rub		
	Latex gloves	-	
	Guidelines for standard precautions		
(d) Diagnostic capacity	Haemoglobin	$n / 8 \times 100$ where <i>n</i> is the total	
	Blood glucose	domain	
	Malaria diagnostic capacity	-	
	Urine dipstick - protein	-	
	Urine dipstick - glucose	-	
	HIV diagnostic capacity	-	
	Syphilis RDT		
	Urine pregnancy test		
(e) Essential medicines	Amlodipine tablet or alternative calcium channel blocker	$n / 25 \times 100$ where n is the total	
	Amoxicillin (syrup/suspension or dispersible tablets)	domain	

 TABLE 1.9.9: GENERAL SERVICE READINESS ITEMS AND INDEX

1. Overview

General service domains	Tracer items	Domain score (mean availability of items)
	amoxicillin tablet	
	Ampicillin powder for injection	
	Aspirin (capsules/tablets)	
	Beclometasone inhaler	
	 Beta blocker (e.g.bisoprolol, metaprolol, carvedilol, atenolol) 	
	Carbamazepine tablet	
	Ceftriaxone injection	
	Diazepam injection	
	 Enalapril tablet or alternative ACE inhibitor (e.g. lisonopril, Ramipril, perindopril) 	
	Fluoxetine tablet	
	Gentamicin injection	
	Glibenclamide tablet	
	Haloperidol tablet	
	Insulin regular injection	
	Magnesium sulfate injectable	
	Metformin tablet	
	 Omeprazole tablet or alternative (e.g. pantoprazole, rabeprazole) 	
	Oral rehydration solution	
	Oxytocin injection	
	Salbutamol inhaler	
	 Simvastatin tablet or other statin (e.g. atorvastatin, pravastatin, fluvastatin) 	
	Thiazide (e.g. hydrochlorothiazide)	
	• Zinc sulphate (tablet or syrup)	
General service readines	(Mean score of the five domains) (a + b + c + d + e) / 5	

Required data source

Facility assessment information is needed to calculate general service readiness; the source for this information is the SARA survey.

1.10 Data archiving

Data archiving includes the acquisition, preservation, documentation, cataloguing and dissemination of microdata.⁸ Archives are useful for promoting research and instruction in the social sciences; ensuring the continued viability and usability of microdata in the future; and providing equitable access to these data within the framework of the national legislation in the interest of all citizens, by protecting confidentiality and following international recommendations and good practices.

Fully documenting and archiving data sets helps ensure that important survey data and metadata are preserved for future reference and analysis. The data documentation, or metadata, helps researchers and other audiences to find the data, understand what the data are measuring and assess the quality of the data.

- **Finding** the data. Names, abstracts, keywords and other important metadata elements help individuals and organizations locate the data sets and variables that meet their needs.
- **Understanding** what the data are measuring and how the data have been created. Descriptions of the survey design and the methods used when collecting and processing the data, allow users to fully comprehend the context of the data.
- Assessing the quality of the data. Information about the data collection standards, as well as any deviations from the planned standards, is important for gauging whether the data are useful for specific uses.

1.10.1 Elements of data documentation

There are three main types of material that constitute ideal documentation for a data set: explanatory material, contextual information and cataloguing material. This represents the minimum to create and preserve a data set, and can be described as the material required to ensure the long-term viability and functionality of a data set. Full understanding of the data set and its contents cannot be achieved without this material.

Explanatory material

Information about the data collection methods

This information describes the data collection process, whether it is a survey; the collection of administrative information; or the transcription of a document source. It should describe the questionnaires used, the methods employed and how these were developed. If applicable, details of the sampling design and sampling frames should be included. It is also useful to include information on any monitoring process undertaken during the data collection as well as details of quality controls.

Information about the structure of the data set

Key to this type of information is a detailed document describing the structure of the data set and including information about relationships between individual files or records within the study. It should include, for example, key variables required for unique identification of subjects across files. It should also include the number of cases and variables in each file and the number of files in the data set. For relational models, a diagram showing the structure and the relations between the records and elements of the data set should be constructed.

⁸ Microdata refers to data on the characteristics of units of a population, such as individuals, households, facilities, or establishments, collected by census, survey or experiment.

Technical information

This information relates to the technical framework and should include:

- the computer system used to generate the files
- the software packages with which the files were created
- the medium on which the data were stored
- a complete list of all data files present in the data set.

Variables, values, coding, and classification schemes

The documentation should contain a full list describing all variables (or fields) in the data set, including a complete explanation and full details about the coding and classifications used for the information allocated to those fields. It is especially important to have blank and missing fields explained and accounted for. It is helpful to identify variables to which standard coding classifications apply, and to record the version of the classification scheme used – preferably with a bibliographic reference to that code.

Information about derived variables

Many data producers derive new variables from original data. This may be as simple as grouping raw age data (age in years) according to groups of years appropriate for the needs of the survey, or it may be much more complex and require the use of sophisticated algorithms. When grouped or derived variables are created, it is important that the logic for the grouping or derivation be clear. Simple grouping, such as for age, can be included within the data dictionary. More complex derivations require other means of recording the information. The best method of describing these is by using flow charts or accurate Boolean statements. It is recommended that sufficient supporting information be provided to allow an easy link between the core variables used and the resultant, derived variables. It is also recommended that the computer algorithms used to create the derivations be saved together with information about the software.

Weighting

The weighting of variables needs to be fully documented, explaining the construction of the variables with a clear indication of the circumstances in which weights should be used. This is particularly important when different weights need to be applied for different purposes.

Data source

Details about the source the data is derived from should be included. For example, when the data source is made up of responses to survey questionnaires, each question should be carefully recorded in the documentation. Ideally, the text will include a reference to the generated variable(s). It is also useful to explain the conditions under which a question would be asked to a respondent, including if possible, the cases to which it applies, and ideally, a summary of response statistics.

Confidentiality and anonymization

It is important to note if the data contain any confidential information on individuals, households, organizations or institutions. Whenever this occurs, it is recommended to record such information together with any agreement on how to use the data, for example, with survey respondents. Issues of confidentiality may restrict the analyses to be undertaken or the results to be published, particularly if the data are to be made available for secondary use. If the data were anonymized to prevent subjects' identification, it is recommended to record the anonymization procedure and its impact on the data, as such modification may restrict subsequent analysis.

Contextual information

Contextual information provides users with material about the context in which the data were collected, and how data were put to use. This type of information adds richness and depth to the documentation, and enables the secondary user to fully understand the background and processes behind the data collection exercise. This also forms a vital historical record for future researchers.

Description of the originating project

Details should be provided about the history of the project or about the process that gave rise to the data set. This should offer information on the intellectual and substantive framework. For example, the description could cover topics such as:

- why the data collection was felt necessary
- the aims and objectives of the project
- who or what was being studied
- the geographical and temporal coverage
- publications or policy developments it contributed to or that arose as a response
- any other relevant information.

Provenance of the data set

Information on the origin of the data set relates to aspects such as the history of the data collection process, changes and developments that occurred in the data themselves and the methodology, or any adjustments made. The following can also be provided:

- details of data errors
- problems encountered in the process of data collection, data entry, data checking and cleaning
- conversion to a different software or operating system
- bibliographic references to reports or publications that stem from the study
- any other useful information on the life-cycle of the data set.

Serial and time-series data sets, new editions

For repeated cross-sectional, panel or time-series data sets, it is helpful to obtain additional information describing, for example, changes in the question text, variable labelling or sampling procedures.

Cataloguing material

Cataloguing material serves two purposes. First, it serves as a bibliographic record of the data set. This allows for the data set to be properly acknowledged and cited in publications, and for the material to act as a formal record for long preservation purposes. Second, it is the basic instrument used for resource discovery, allowing the data set to be uniquely identified within the collection by providing appropriate information to help secondary users identify the study as being useful to their purpose.

1.10.2 Metadata standards

Traditionally, data producers and archivists produced expansive, text-based codebooks. Today, various metadata alternatives, such the Data Documentation Initiative (DDI) and the Dublin Core Metadata Initiative (DCMI), have been developed for the documentation and cataloguing of microdata and related materials

according to international standards. These new type of 'codebooks' are based on Extensible Markup Language (XML), a type of regular text file that tags for meaning – rather than appearance – and can be viewed and edited using any standard text editor. XML files can be searched and queried like a regular database and can be edited.

Data Documentation Initiative (DDI)

The Data Documentation Initiative (DDI) is an effort to establish an international XML-based standard for microdata documentation. Its aim is to provide a straightforward means to record and communicate to others all the salient characteristics of microdata sets. By creating a consistent framework for microdata documentation, the DDI has several key features: interoperability, richer content, multi-purpose documentation, online analytical capability and search capability.

The DDI elements are organized in five sections.

Section 1.0. Document description

A study (survey, census or other) is not always documented and disseminated by the same agency as the one that produced the data. It is therefore important to provide information (metadata) not only on the study itself, but also On the documentation process. The document description consists of overview information describing the DDI-compliant XML document, or, in other words, "metadata about the metadata".

Section 2.0. Study description

The study description consists of overview information about the study. This section includes information about how the study should be cited; who collected, compiled and distributes the data; a summary (abstract) of the content of the data; and information on data collection methods and processing.

Section 3.0. Data file description

This section is used to describe each data file in terms of content; record and variable counts; version; producer; and so on.

Section 4.0. Variable description

This section presents detailed information on each variable, including literal question text; universe, variable and value labels; and derivation and imputation methods.

Section 5.0. Other material

This section allows for the description of other materials related to the study. These can include resources such as documents (e.g. questionnaires, coding information, technical and analytical reports, interviewer's manuals), data processing and analysis programs, photos and maps. However, the DCMI (described below) provides a standard for documenting digital resources such as questionnaires and reports.

Dublin Core Metadata Initiative (DCMI)

The Dublin Core Metadata Initiative (DCMI) is an open forum to develop the Dublin Core metadata standard, which is a simple set of elements for describing digital resources. This standard is particularly useful to describe resources related to microdata such as questionnaires, reports, manuals, data processing scripts and programs. A major reason behind the success of the Dublin Core metadata standard is its simplicity. From the outset it has been the goal of the designers to keep the element set as small and simple as possible to allow the standard to be used by non-specialists. In its simplest form the Dublin Core consists of 15 metadata elements, all of which are optional and repeatable. The 15 elements are:

- 1. title
- 2. subject (topic)
- 3. description: an abstract, a table of contents, or a free-text account of the content
- 4. type: the nature or genre of the content of the resource
- 5. source
- 6. relation: a reference to a related resource (rarely used)
- 7. coverage: the extent or scope of the content of the resource (e.g. spatial location or time period)
- 8. creator
- 9. publisher
- 10. contributor
- 11. rights: a rights management statement for the resource
- 12. date
- 13. format
- 14. identifier
- 15. language.

1.10.3 Creating metadata for SARA

Metadata can be created through a multitude of media including simple word processing programs and software application programs. This section provides guidance on creating metadata for SARA by identifying key elements to be included and by providing information on tools available to assist in creation of metadata.

Required elements

When creating a metadata document using a simple word processing program, the following elements need to be included. Much of this information will have been generated as part of the data processing steps.

Survey description

DOCUMENT DESCRIPTION

The document description serves as an introduction to the metadata as a whole. It provides background information such as the study title, document producer(s), date of production and version number.

STUDY DESCRIPTION

The study description serves to identify the study itself and to provide overview information, as well as the project scope, coverage and sampling, and information on data collection, editing, appraisal and access. This section also names producers and sponsors, and describes points of contact, and disclaimers and copyrights.

Data set(s)

FILE DESCRIPTION

The file description of a data set provides the data set contents, its producer and the version. It should also include an explanation of how missing data are coded or accounted for, as well as any other relevant notes. When applicable, a section on processing checks should be included. This element serves to provide information about the types of checks and operations that have been performed on the data file to make sure that the data are as correct as possible, e.g. consistency checking.

VARIABLES

The variables section of an archive consists of detailed descriptions of the actual data.

The *variables list* is typically a table listing every variable in the data set and providing for each the variable number, name and label. This list also provides the literal question associated with the variable, the variable format (character or numeric, number of units), and the number of valid and invalid cases (see Table 1.10.3).

TABLE 1.10.1: VARIABLES LIST

#	Name	Label	Туре	Format	Valid	Invalid	Question
1	V_001	Facility	Discrete	Character-12	97	0	Record the name
		Name					of the facility

The *variables description* is more detailed than the variable list. It includes variable information (type, format, missing value coding), statistics (valid and invalid), literal question, and any notes (see Table 1.10.4).

TABLE 1.10.2: VARIABLES DESCRIPTION

#1 V_001: Facility name			
Information	[Type= discrete] [Format=character] [Missing=*]		
Statistics	[Valid=97 /-] [Invalid=0 /-]		
Literal question	Record the name of the facility		
Notes			

External resources

TYPES OF RESOURCES

External resources encompass all of the documents contributing to the implementation of the survey or stemming from the results of the survey. Examples include:

- questionnaires
- reports
- databases
- photos, videos, etc.
- maps or geospatial data
- technical documents
- analytical or administrative documents.

RESOURCE INFORMATION

Each external resource should be accompanied by relevant descriptive information.

Identification

- type of resource
- title
- authors: the individuals or organization primarily responsible for creating the resource
- date: the date on which the resource was created or last modified
- country: all countries within the scope of a resource
- language
- format
- an ID number, if applicable: an unambiguous reference to the resource.

Contributor and rights

- contributor(s): individuals or organizations who have supported or contributed to the development of the resource (including funding agencies)
- publisher(s): individuals or organizations responsible for disseminating the resource
- rights: a clear and complete description of the usage rights, if relevant.

<u>Content</u>

- description: an account of the content of the resource
- abstract
- table of contents: a listing of all sections of the resource
- subjects: key topics discussed in the resource.

Available tools

The Microdata Management Toolkit⁹ developed by the World Bank Data Group is designed to address the technical issues facing data producers. It provides one of the most straightforward ways to create comprehensive metadata that adhere to international standards. The aim in developing the Toolkit was to promote the adoption of standards for international microdata documentation, dissemination and preservation, as well as to foster best practices by data producers in developing countries.

The Toolkit consists of:

- a Metadata Editor, which documents data in accordance with international standards;
- an International Household Survey Network (IHSN) Report Center, which generates metadata reports from inputs into the Metadata Editor;
- an Explorer, which allows users to view metadata and to re-export data in common formats;
- a CD-Rom Builder, which generates user-friendly outputs (CD-ROM, web) for dissemination and archiving.

Templates for SARA survey archiving are publicly available through the IHFAN web site at <u>http://www.ihfan.org/home/index.php?editable=no&page type=catalog</u>.

⁹ The Microdata Management Toolkit is free and available for download along with a user manual at: <u>http://www.ihsn.org/home/index.php?q=tools/toolkit</u>

1.10.4 Data archiving

Today, data archives are most always digital and are ideally web-based or are made publicly available through the Internet. While this can be accomplished through many different types of media, SARA makes use of the National Data Archive (NADA) which is a free, standardized application for publishing data archives.¹⁰

National Data Archive (NADA) tool

The International Household Survey Network (IHSN) developed the national data archive (NADA) as a complement to the Microdata Management Toolkit. NADA is a web-based survey cataloguing system that serves as a portal for researchers to browse, search, apply for access, and download relevant census or survey data and metadata.

NADA makes use of the XML-based international standards such as the DDI and Dublin Core and is a powerful instrument that facilitates the process of releasing study metadata and microdata to the user community. NADA is a tool for informing users about the existence and characteristics of survey, census or other microdata sets, and for sharing metadata and (optional) disseminating microdata files. NADA does not provide tools for data tabulation or analysis. It aims to provide users with detailed and searchable documentation of microdata sets, along with information on policies and procedures for their access and use. NADA comes as a prepackaged but fully customizable web site. At the core of NADA is the data catalogue, which:

- provides summary information on each survey;
- provides access to reports, tables and other analytical output;
- provides data access policies to the user community and facilitates access by serving as an implementing tool of the data access policy;
- provides links to related survey metadata;
- facilitates searches at the variable level and displays variable-level information;
- provides authorized users with access to the data (via direct access or through online forms), with conditions for access clearly stated;
- keeps a log of user requests;
- links to the HTML output as provided by the CD-ROM Builder of the Microdata Management Toolkit;
- includes an automatically-generated history of added/updated data sets via an RSS feed;
- is easy to maintain and use.

The data catalogue interface is interactive, allowing users to sort and search the catalogue by study elements and/or data variables, or find out detailed information through the survey's metadata.

WHO has created a national data archive for SARA surveys, which can be located at http://apps.who.int/healthinfo/systems/datacatalog/index.php/catalog. This site serves as an example of how a data archive can be created using the NADA software.

¹⁰ NADA is available to download free of charge at: <u>http://www.ihsn.org/home/index.php?q=tools/nada</u>.

References

1. *International Health Partnership and related initiatives (IHP+)*. Geneva, World Health Organization and Washington DC, The World Bank (<u>http://www.internationalhealthpartnership.net/en/home</u>, accessed 17 December 2011).

2. *Monitoring, evaluation and review of national health strategies. A country-led platform for information and accountability.* Geneva, World Health Organization, 2011.

3. *Service availability mapping (SAM)*. Geneva, World Health Organization (<u>http://www.who.int/healthinfo/systems/samintro/en/index.html</u>, accessed 17 December 2011).

4. *Service provision assessment (SPA) overview*. Maryland, MEASURE DHS, ICF International (<u>http://www.measuredhs.com/aboutsurveys/spa/start.cfm</u>, accessed 17 December 2011).

5. *Measuring medicine prices, availability, affordability and price components,* 2nd ed. Geneva, World Health Organization and Health Action International, 2008 (http://www.haiweb.org/medicineprices/manual/documents.html and http://www.who.int/medicines/areas/access/medicines_prices08/en/, accessed 17 December 2011).

6. *Monitoring the building blocks of health systems: a handbook of indicators and their measurement strategies.* Geneva, World Health Organization, 2010 (<u>http://www.who.int/healthinfo/systems/monitoring/en/index.htm</u>, accessed 17 December 2011).

7. Creating a master facility list. Draft document. Geneva, World Health Organization, 2012

8. Data Quality Review (DQR): A toolkit for facility data quality assessment. Working document. Geneva, World Health Organization, 2015

- 9. Health workforce target reference
- 10. Outpatient service utilization target reference
- 11. Inpatient service utilization target reference