Data Management Competency Framework
Data Management Competency Framework
Contents

Foreword iv
Acknowledgements v
Abbreviations vi
Executive summary vii

1. Introduction 1
   Background ........................................... 1
   Objective ............................................. 2
   Target audiences .................................... 2
   Framework structure and components .................. 2

2. The Competency Framework 4
   Competency Area 1: Data Generation .................. 6
   Competency Area 2: Data Processing .................... 16
   Competency Area 3: Data Analysis .................... 28
   Competency Area 4: Data Usage ....................... 40
   Attitudes ............................................. 50

3. Implementation in the field 52
   Case story: Papua New Guinea implementation ...... 55

References 56
Annexes 58
The needs of data-driven global and national strategies call for a rethinking of the way data are managed and how we use data with purpose. New technologies have accelerated growth in data types, volumes and complexity. Complicated and emerging health challenges that impact both social and economic environments simultaneously require maximizing the use of data for strategic dialogue and decision-making.

The health information management profession has been making a vital contribution to supporting health care and health-care delivery. With new demands comes the requirement for the health information workforce (HIW) to stay abreast of these changes in terms of what they mean for their roles. However, there remains a significant divergence across the globe in the levels of HIW education, training and development environments and the levels of professionalization, resulting in wide variations and differences in the levels of work readiness among graduates preparing to enter the workforce.

The World Health Organization (WHO) Regional Office for the Western Pacific, with support from WHO country offices and headquarters and in collaboration with partner organizations, is pleased to present the Data Management Competency Framework, a tool that provides a clear definition of each component of the data life cycle as well as the required skills and knowledge for different HIW levels, thus providing the capability to identify current competency gaps, measure competency development and identify future competency needs to support Member States in promoting sustainable and integrated capacity-building for both the short and longer terms.

The Framework is a comprehensive, practical guide that Member States can adapt to suit their own circumstances and their own vision and goals. Its publication is very timely, coinciding with the requirement to align the changing environment with the need for utilization of multiple data resources. This is a period, especially due to the coronavirus disease (COVID-19) pandemic, when all health systems face stringent health challenges, greater demands for efficiencies and higher expectations of health information.

This Framework expertly demonstrates how to take action now to achieve strategic, sustainable and integrated capacity-building for all HIW levels and types of health institutions in a tailored way. While it brings decision-makers, managers, implementors and frontline health workers at national and subnational levels much closer together, the Framework also encourages and provides the opportunity for active multi-stakeholder collaboration, including both horizontal approaches between different departments and their human resource departments, as well as vertical approaches across national, provincial, district and health-centre levels.

Dr Kidong Park, M.D., Ph.D.
Director
Data, Strategy and Innovation (DSI)
WHO Regional Office for the Western Pacific
Acknowledgements

The Data Management Competency Framework was developed under the guidance of Dr Kidong Park, Director of the Data, Strategy and Innovation (DSI) Group at the World Health Organization (WHO) Regional Office for the Western Pacific, and Dr Gao Jun, Team Coordinator, Health Information and Intelligence unit (HII) of DSI.

The core writing group members consisted of Mengjuan Duan, Eugene O’Curry and Robert Arciaga from HII.

Technical inputs at the conceptualization stage were provided by Brian Riley, Vladimir Choi, Roberta Pastore, Priya Mannava, Linh-Vi Le, Fukushi Morishita and Elena Dolmat from the WHO Regional Office for the Western Pacific.

Further technical inputs were provided at the drafting stage by Ana Mendez-Lopez, Tracy Yuen, Linh-Vi Le, Fukushi Morishita and Josaia Tiko from the WHO Regional Office for the Western Pacific; Roland Dilipkumar Hensman, Phatsaline Vongsaly, Carl Massonneau, Hiroko Henker, Daisuke Asai, Khamphao Keonantatilard and Vilath Seevisay from the WHO country office in the Lao People’s Democratic Republic; and Hong Anh Chu, Doris Ma Fat, Nenad Kostanjsek, Nelly Biondi, Luhua Zhao, Bochen Cao and Derrick Muneene from WHO headquarters.

Contributions to the Papua New Guinea case story were provided by Sevil Huseynova, Priya Mannava, Anna Maalsen and Getinet Adenager from the WHO country office in Papua New Guinea and Manah Dindi from the National Department of Health, Papua New Guinea.

The draft Data Management Competency Framework was reviewed by the following external technical experts: Vicki Bennett, Filippa Pretty, Nicola Richards, Georgia Savvopoulos, Ashna Kumar, Kate Spyby, Brooke Macpherson, Breanna Harnett, Miriam Lum On and Ashleigh Bennett from the Australia Institute of Health and Welfare; Willy Chan from the New South Wales Local Health District; Cameron Barnes from St Vincent’s Health Australia; Sheree Lloyd from the Griffith University and University of Tasmania; and Louise Edmonds from Calvary.
Abbreviations

ANOVA analysis of variance (statistical technique)
COVID-19 coronavirus disease
CAPI computer-assisted personal interview
DHIS2 District Health Information System 2
DBMS database management system
dsi Data, Strategy and Innovation (Group)
ELT extract, load and transform
ETL extract, transform and load
FHIR Fast Healthcare Interoperability Resources
GIS geographic information system
HIW health information workforce
HIS health information system
HR human resource
INLA integrated nested Laplace approximation
ISO International Organization for Standardization
KTA Knowledge to Action
MDR-TB multidrug-resistant tuberculosis
M&E monitoring and evaluation
R (programme scripting language)
RHIS routine health information system
SMART specific, measurable, achievable, relevant, time-bound
SOP standard operating procedure
SAS statistical analysis software
SDGs Sustainable Development Goals
SPSS Statistical Package for the Social Sciences
SQL structured query language
WHO World Health Organization
Executive Summary

The effective use of data in the management and delivery of public health services has long been understood as critical by public health professionals. Over recent decades, countries have made significant investments in the improvement of data generation and quality, but mostly from the perspective of data generation rather than that of data users. New technologies have accelerated growth in data types, volumes and complexity. Emerging and complicated health challenges that impact both social and economic environments simultaneously require maximizing the use of data for strategic dialogue and decision-making.

With new demands comes the requirement for the health information workforce (HIW) to stay abreast of these changes in terms of what they mean for their roles. Developing the range of human competencies necessary to understand and interpret the ever-changing risks to and complexities of public health through the effective use of data management skills remains an ongoing challenge in many countries. There remains a significant divergence across the globe in the levels of HIW education, training and development environments and the levels of employee professionalization, resulting in wide variations and differences in the levels of work readiness among graduates preparing to enter the workforce. These changing needs and expectations call for a rethinking of the way data are managed and how we use data with purpose.

To address this challenge, the Data, Strategy and Innovation (DSI) Data Group was set up to lead office-wide collaboration on key health information priorities with "data focal points" across all divisions of the World Health Organization (WHO) Regional Office for the Western Pacific. In collaboration with a team of WHO colleagues from headquarters and country offices, external partners and experts, the Data Group has designed and developed the Data Management Competency Framework as a practical tool to provide both a structure and methodology to enable HIW employees (who have cause and need to engage with health data), their line managers and human resource (HR) managers to define the competencies required for the identified data management roles within their organizations. The value and benefits deriving from the use of the Framework are numerous and varied. They range from identifying individual, unit, department and divisional competency gaps and training needs at an operational level, to informing and guiding employee recruitment and selection, training and development, and team formation at an organizational level.

The Framework is designed around the four core stages of the data management cycle: data generation, processing, analysis and usage. Within the Framework, these stages are classified as competency areas, and within these four competency areas, a further
17 competency domains have been identified. These domains are effectively sets of sub-competencies within the wider competency areas. The Framework provides users with detailed domain definitions and specific descriptors of the technical knowledge and skill sets covered within each domain. The Framework further provides and describes four proficiency levels, i.e., levels of competence within each domain, and describes the technical knowledge and skills that determine the achievement of each proficiency level. This is rounded out by attitudinal domains that describe behaviours and mindsets that are essential for long-term sustainable capacity-building.

The Framework provides in-country management, for the first time, with a detailed, comprehensive, integrated and coherent tool to define and assess current and future competency needs for their HIW. The objective in producing and introducing this Framework is to make it the go-to competency mapping and development tool for health information employees, their line managers and HR managers across the WHO Western Pacific Region. Using the Framework comprehensively and consistently will yield a host of benefits for the HIW.

As with all organizational change initiatives, the key to success is effective implementation. Local ownership of and responsibility for implementation are critical. Country management must embrace the concept and put in place the necessary oversight and implementation structures to ensure successful adoption. This document describes a seven-step process to guide those entrusted with responsibility for implementation.
1. Introduction

Background

Accurate, reliable and timely data are critical for driving global and national health strategies and for ensuring the advancement of national health goals in Member States of the World Health Organization (WHO) Western Pacific Region, while simultaneously meeting internationally agreed commitments such as the health-related Sustainable Development Goals (SDGs). Over recent decades, countries have made significant investments in the improvement of their health information systems (HIS) to respond to the ever-increasing demands for new and better data. Due to the coronavirus disease (COVID-19) pandemic, decision-makers and broader populations have recognized the importance of resilient and robust HIS, with a trend of increased investments over the past two years (1). Multiple digital tools and platforms have been developed or embedded into current systems in Member States (2).

A common challenge, however, is that health information workforces (HIWs) do not have the robust capacities to participate in data generation, processing, analysis and usage in a reliable, consistent and timely manner, which minimizes the impact of investments made in HIS development.

The development of HIW capacity is lagging behind the continuous improvement of HIS, especially since the COVID-19 pandemic. There are large variances in the roles, responsibilities and competencies across the HIW, both within and between countries (3). Capacity-building initiatives and activities normally follow traditional, fragmented and ad hoc approaches (Fig. 1), which regularly result in suboptimal and non-sustainable investment returns.

![Fig. 1. Common challenges for traditional capacity-building](image-url)

- Driven by donors or specific technical programmes to fulfill short-term and ad-hoc needs. Hence the lack of sustainable and systemic plan.
- Lack of connection with wider organizational missions and objectives
- Lack of cross-divisional cooperation which results in suboptimal resource mobilization for identification of potential duplications and discovery of vulnerable areas
As a first step in addressing these systemic weaknesses, the WHO Regional Office for the Western Pacific created a cross-functional, multidisciplinary Data Group to devise and implement an overarching strategic response to this problem. Within this group of specialists, a Capacity-Building Working Group was formed to lead the design and development of a systematic and integrated approach to strengthening the HIW capacity of Member States at different levels and types of health institutions. The outcome of these efforts, in collaboration with a mixed team of WHO colleagues from headquarters and country offices, external partners and experts, is this Data Management Competency Framework, which is described in more detail on the following pages.

**Objective**

The Framework is an essential tool for health information management teams in Member States to identify capacity gaps and plan and drive the competency development of their HIW across different levels of their health institutions and organizations. More specifically, it is a reference tool for activities that can:

- measure the skill and knowledge gaps of individuals or homogeneous role types within the HIW at each step of the data life cycle (Fig. 2);
- identify and specify the required skills and knowledge within the workforce to match current roles and responsibilities;
- identify and map the priority areas for both short and long-term training and development planning at the individual, unit, department and division level within the HIW in the organization; and
- inform, guide and support strategic human resources (HR) management activities such as recruitment, selection and appointments, team formation and composition based on the mix and balance of available competencies.

**Target audiences**

The principal target audiences for the Framework are the line and HR management teams from Member States to enable them to create data management competency profiles (required and actual) for their health information employees. Such roles would typically include but are not limited to:

- decision-makers, managers, implementors and front-line health workers who develop, manage and use HIS, such as health and clinical information professionals and officers;
- decision-makers, managers, implementors and front-line health workers who input data and extract information for routine work, such as specific disease-control managers, records managers, surveillance and monitoring officers, and data analysts; and
- adoption/implementation leaders and influencers, such as national and subnational health system decision-makers, front-line functional line managers, HR managers and training managers.

Software designers and information technology specialists are not included.

**Framework structure and components**

The Framework (illustrated in Fig. 2) is designed around four competency areas or “pillars” that are based on the four stages of the data management cycle: data generation, processing, analysis and usage. Within each competency area are competency domains (17 domains in total), which are distinct and essential sub-competencies. Within each competency domain, there are three competency dimensions (subject knowledge, practical skills and personal attitude), which, when combined, determine the overall level of competence of an individual employee.
Knowledge and skills are further categorized into four proficiency levels (ranging from learner-beginner to master practitioner), which describe the varying degrees of proficiency achievable for each domain. Attitudes are further categorized into attitudinal domains which collectively describe behaviours and mindsets which create an environment of excellence and respect leading to long-term sustainable capacity-building. Details can be found in the following chapter.
2. The Competency Framework

- **Data Generation**
- **Data Processing**
- **Data Usage**
- **Data Analysis**

Knowledge, Skills, Attitude
The Data Management Competency Framework is a practical tool to enable data management practitioners (that is line managers and HR/training managers):

- to identify required data management role competencies and proficiency levels (required competency profile);
- to identify employees’ current data management competencies and proficiency levels (actual competency profile);
- to map the gaps that exist between required and actual competency profiles; and
- to draft competency development plans at individual, team, department, division and organizational levels.

The Framework is structured around four competency areas or "pillars" reflecting the four distinct stages of the data management cycle. Within each competency area are competency domains (17 in total), which are a set of distinct sub-competencies for each competency area. Within each domain, the Framework sets out the three competency dimensions (subject knowledge, practical skills and personal attitudes), which are the determinants of overall competence.

Each domain contains four designated proficiency levels (ranging from learner-beginner to master practitioner), which describe the incremental levels of proficiency by reference to a set of described technical–professional tasks that are specific, measurable and role relevant and which will determine the level of proficiency achieved by an individual employee.

![Fig. 2. Structure of the Data Management Competency Framework](image-url)
Competency Area 1
Data Generation

- Data management planning
- Data creation
- Data collection
- Data maintenance
## Competency domains

<table>
<thead>
<tr>
<th>#</th>
<th>DOMAINS</th>
<th>DEFINITIONS/EXPLANATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data management planning</td>
<td>Data management planning is the process of delivering a formal written document (data management plan) that describes the process and methodology by which data will be created, collected and maintained throughout the life cycle of a proposed survey, trial or research activity and beyond (4-6). Factors of relevance and concern include data source identification and validation, data quality assurance and control, selection of data types (primary, secondary, tertiary), appropriate collection methodology and data maintenance (see Domain 4 below).</td>
</tr>
<tr>
<td>2</td>
<td>Data creation</td>
<td>Data creation is the process of bringing data into the realm of (human) access in an organized format for collection, analysis and interpretation (7). This can be achieved by means of quantitative and/or qualitative processes (for example, questionnaires, observational studies) and approved methods of academic and scientific enquiry and examination. In HIS, this step is normally called recording using a registry (for example, facility logbook, patient form, community health logbook) and can be done in the form of a manually prepared document or captured from the source using a data-capture device, such as a barcode reader.</td>
</tr>
<tr>
<td>3</td>
<td>Data collection (including sourcing)</td>
<td>Data collection is the process of gathering data of interest in an established systematic fashion that enables answers to be found to stated questions, to test hypotheses and to evaluate outcomes (8). Data sourcing is the process of identifying, extracting and using data from different (often multiple) internal and/or external sources to serve specific purposes (9).</td>
</tr>
<tr>
<td>4</td>
<td>Data maintenance</td>
<td>Data maintenance is concerned with setting standards for how data will be gathered, organized, stored and curated post capture and with ensuring these standards are met consistently, regardless of whether the storage is paper based or electronic (10). Data maintenance covers aspects such as organizing, coding, storing, preserving, archiving and sharing, and raises inherent issues such as data sensitivity, security protection and data integrity.</td>
</tr>
</tbody>
</table>
## Domain 1: Data management planning

### Learner-Beginner

**KNOWLEDGE DIMENSION**

**Be able to describe and explain:**

- What a data management plan (or adapted term for localized use) is
- The purpose, role and importance of a data management plan in the data-generation process
- Why data are organizational assets and why data security risks should be monitored and mitigated
- The critical limitations of data
- The core International Organization for Standardization (ISO) data standards and data management planning principles
- The most common risks associated with data
- How and under what circumstances identified data risks should be escalated through the risk management structure

**SKILLS DIMENSION**

**Be able to do the following:**

- Identify any missing key information in a data request (for example, target population) and know when to seek further clarification
- Identify where ISO data standards and data management planning principles have been applied (or not) in a sample data management plan
- Categorize the levels of potential value and utilization of different data types and datasets based on the identified limitations of the data provided
- Recognize and categorize different data types and datasets in ascending order of value to the organization (D/C/B/A – low, medium, high, mission essential)
- Categorize different data types and datasets in ascending order of risk to the organization (D/C/B/A – low, medium, high, mission threatening)

### Novice Practitioner

**KNOWLEDGE DIMENSION**

In addition to the Learner-Beginner knowledge, **be able to describe and explain:**

- The factors that influence and determine the design of data collection instruments and data presentation tools and techniques (for example, the audience, the nature of the data, whether to store in a structured or unstructured manner)
- Inherent risks within the overall data collection process
- The expected contents, structure and core elements of a data management plan (for example, the what and the how of data collection, application of the data management cycle)
- The organizational process of escalating identified data security risks

**SKILLS DIMENSION**

In addition to the Learner-Beginner skills, **be able to do the following:**

- Draft a basic data management plan based on a provided template
- Identify data risks in a provided data management plan and draft a basic risk removal/mitigation plan to address them
- Design basic data collection instruments and presentation tools and techniques
- Initiate an appropriate data security risk escalation process
Domain 1: Data management planning (contd)

**Independent Practitioner**

**KNOWLEDGE DIMENSION**

In addition to the Novice Practitioner knowledge, **be able to describe and explain**:

- Who should be involved and what technical inputs are required in drafting, reviewing and approving the data management plan including details and specifics
- The other three data generation domains with individuals specialized in these domains to create a cohesive plan
- How to map hardware, software and HR plans for a data collection exercise
- How the plan fits within the broader institutional data management vision and/or strategy
- The identified data risks and the mitigating actions in the organizational context

**SKILLS DIMENSION**

In addition to the Novice Practitioner skills, **be able to do the following**:

- Set realistic goals and timelines in a data management plan
- Develop a comprehensive and realistic data management plan based on a HIS question or data demand
- Develop indicators to monitor progress as per the data management plan
- Adapt international indicators to national routine health information system (RHIS) contexts

**Master Practitioner**

**KNOWLEDGE DIMENSION**

In addition to the Independent Practitioner knowledge, **be able to describe and explain**:

- The existing data management plans in the organization and how they relate to each other
- The criteria for evaluating the effectiveness of the plan
- The data resources already available and data gaps
- The process for monitoring and evaluating the effectiveness of a data management plan
- Leadership, education and dissemination, including inputs in the areas of governance, policy and procedures design, and quality assurance of data risks
- The contingency planning process and the plan for continued data generation (for example, changing the focus of the data need and existing data generation mechanisms) in the event of a force majeure or catastrophic incident
- The policy and mechanisms for ensuring effective data governance
- How to minimize risks associated with data creation, collection and maintenance

**SKILLS DIMENSION**

In addition to the Independent Practitioner skills, **be able to do the following**:

- Foresee and anticipate future data needs
- Demonstrate leadership by championing and advocating for data needs and projects to gain high-level support and commitment during the data management planning process
- Critique an existing data management plan by clearly identifying areas of concern and provide solutions
- Identify and specify the resources necessary for a data management plan
- Identify roadblocks to successfully fulfilling the data request (for example, budget and HR constraints, data request too complex), and suggest alternative solutions to meet the needs of the data request
# Domain 2: Data creation

## Learner-Beginner

### KNOWLEDGE DIMENSION

**Be able to describe and explain:**

- The differences between input, output, outcome and impact health indicators
- The concept of bias, how bias can occur and its potential effects on the dataset
- The basics of research design: types of primary research studies (for example, observational, experimental) and the key elements of each one

### SKILLS DIMENSION

**Be able to do the following:**

- Determine which data creation method(s) is/are the most appropriate for the data request based on the pros and cons of each in context
- Appreciate different data creation methods and their suitability for different data requests

## Novice-Practitioner

### KNOWLEDGE DIMENSION

In addition to the Learner-Beginner knowledge, **be able to describe and explain:**

- The distinctions between survey, RHIS and research data generation
- The distinct types of data (longitudinal, event, aggregate) and their use
- The different data creation methods and their application contexts for analysis and use
- The different data collection methods
- The relative strengths and weaknesses of each data collection method
- Awareness that some groups might be underrepresented in your dataset depending on how collection is set up

### SKILLS DIMENSION

In addition to the Learner-Beginner skills, **be able to do the following:**

- Properly define data elements required in a relevant aggregate report or data request
- Determine and describe the key components of the data creation methods (for example, for prospective data collection, define the target population, sampling method and inclusion criteria) with moderate supervision
- Identify reliable data sources for secondary analysis (for example, national government data sources) and relevant peer-reviewed studies
- Map existing data related to the indicator to measure
- Identify whether the data required already exists and assess the feasibility of obtaining it
COMPETENCY AREA 1: DATA GENERATION

Domain 2: Data creation (contd)

**Independent Practitioner**

**KNOWLEDGE DIMENSION**

In addition to the Novice Practitioner knowledge, **be able to describe and explain:**

- How data needs are identified
- The different systems available for collecting data
- The different data requirements and standards on how data should be generated and collected
- The different sampling methods available when collecting data
- The presence or existence of bias in a particular dataset

**SKILLS DIMENSION**

In addition to the Novice Practitioner skills, **be able to do the following:**

- Produce a realistic timeline for data creation (daily/monthly/quarterly/annual)
- Produce a set of indicators that are specific, measurable, achievable, relevant and time-bound
- Assess the feasibility of using one methodology over another
- Define the indicators’ components to avoid any subjective interpretation of the questions
- Produce clear standard operating procedures (SOPs) for data creation, anticipating cognitive and other potential biases

**Master Practitioner**

**KNOWLEDGE DIMENSION**

In addition to the Independent Practitioner knowledge, **be able to describe and explain:**

- The main data creation activities within the organization
- What data sources are available and how they should be used for specific measurement indicators
- Potential limitations, new or upcoming changes to national and international data standards
- Potential influences and/or political barriers (for example, funding agencies, contention between departments of health units) that may impact data creation
- Situations where using specific data creation methods have failed or succeeded
- The interoperability of data collection systems
- The role and relevance of data creation in support of health policies or to monitor goals in the short and long-term
- Strategies that can be applied to minimize bias and taking a leadership role in defining and institutionalizing these strategies

**SKILLS DIMENSION**

In addition to the Independent Practitioner skills, **be able to do the following:**

- Identify a data need based on the observation and analysis of trends
- Develop a user manual explaining the data creation and how it is linked to all other components such as data processing and analysis
- Communicate an overall view of various data creations within the institution to avoid duplication or to leverage existing data creations
- Pilot and validate the developed tools for data collection
- Communicate the role and relevance of data creation in supporting health policies or in monitoring set goals in the short and long-term
- Prepare a contingency plan if data creation fails or does not achieve the initial goals
## Domain 3: Data collection

### KNOWLEDGE DIMENSION

**Learner-Beginner**

**Be able to describe and explain:**

- The tools for routine data collection and their application in context
- Options in obtaining data and the advantages and disadvantages associated with each option
- The basics of information flow, data standards for data collection and the potential consequences of deviating from data standards

### SKILLS DIMENSION

**Be able to do the following:**

- Encode relevant data accurately and reliably into the RHIS
- Use the digital data collection tool to do basic tasks (for example, view datasets, enter data)
- Identify data requirements and extract relevant, reliable data to suit the purpose and task
- Check, validate and verify the data source(s)

### Novice Practitioner

**KNOWLEDGE DIMENSION**

In addition to the Learner-Beginner knowledge, **be able to describe and explain:**

- How to identify different and reliable data sources
- Methods and standards for generating patient/client identification in digital/paper formats that can be easily authenticated and linked to existing standards, if any
- Commonly used data collection tools, techniques and methods, their advantages and disadvantages and their application contexts

### SKILLS DIMENSION

In addition to the Learner-Beginner skills, **be able to do the following:**

- Make corrections or updates to data based on judgement
- Identify a merging ID or be able to create a merging ID
- Break down the collected information and record it as specified in the data maintenance plan (for example, if collecting data from test results from medical records, what key data points should be collected from the lab reports and how should they be entered to fit the data element specification?) to evaluate the outcomes of the study
- Identify the format of datasets and be able to switch formats (for example, from wide to long) to suit needs and circumstances
- Implement a data collection plan as specified (for example, use their language, cultural, subject-matter knowledge/skills to conduct surveys)
- Analyse gaps in data sources and address data gaps
- Add new data to an existing dataset by creating a new case/observation, appending/binding or merging data from different sources
**DOMAIN 3: DATA COLLECTION**

**INDEPENDENT PRACTITIONER**

**KNOWLEDGE DIMENSION**

In addition to the Novice Practitioner knowledge, be able to describe and explain:

- The relative strengths and weaknesses of each of the collection tools, techniques and methods, including the difference between paper-based and digital tools (and how they can complement each other)
- The relevant data sources for a given data collection activity or programme

**SKILLS DIMENSION**

In addition to the Novice Practitioner skills, be able to do the following:

- Design client coding systems for different data collection exercises
- Design data validation rules for digital tools and clear SOPs
- Map existing information flows related to the data to collect
- Design and develop relevant and assessor-friendly data collection tools for defined indicators
- Select relevant data collection methods for each defined indicator
- Design a set of data validation rules for digital tools
- Train assessors on data collection
- Produce or reuse standards and codes for unit of analysis (for example, client, household)
- Identify situations when the specified method may need to be modified due to context and seek approval from higher authorities, as appropriate
- Confidently launch and monitor new rounds of data collection
- Source comparable data from paper-based (logbook/tally) or digital sources
- Ascertain the data collected are fit for purpose

**MASTER PRACTITIONER**

**KNOWLEDGE DIMENSION**

In addition to the Independent Practitioner knowledge, be able to describe and explain:

- How to design client coding systems for different data collection exercises
- The advanced coding system for data collection, including but not limited to Fast Healthcare Interoperability Resources (FHIR)
- The implied cost of data collection
- Standard requirements for international agencies such as the United States Agency for International Development and Gavi, the Vaccine Alliance
- The limitations of a selected collection system
- The potential issues with the security model of various collection systems

**SKILLS DIMENSION**

In addition to the Independent Practitioner skills, be able to do the following:

- Design and lead a training programme for assessors on data collection
- Design and manage a data collection exercise in the field
- Adopt and adapt different standards in data collection as per specific areas such as international forms
- Organize user-acceptance testing of digital tools for data collection
- Seek or propose the best available market tools for data collection
Domain 4: Data maintenance

**KNOWLEDGE DIMENSION**

Be able to describe and explain:

- What a data maintenance plan is and its role and importance in the data-generation process
- The various and different responsibilities involved with the ownership of data
- Why and how legislation/policy may influence restrictions on data storage
- The various and differing considerations in data maintenance such as data ownership, legislative restrictions on data storage, data integrity, data retrieval, data recovery, data retention, data management responsibility, data disposal and storage cost

**SKILLS DIMENSION**

Be able to do the following:

- Outline the framework, structure or contents of a basic data maintenance plan
- Modify an existing database
- Create a basic database according to a set of given standards

---

**Novice Practitioner**

**KNOWLEDGE DIMENSION**

In addition to the Learner-Beginner knowledge, be able to describe and explain:

- The various considerations in data storage including data ownership, legislative restrictions on data storage, data integrity, data retrieval, data recovery, data retention, data management responsibility, data disposal and storage cost
- The common database software and database management tools used in data maintenance and the pros and cons of each one

**SKILLS DIMENSION**

In addition to the Learner-Beginner skills, be able to do the following:

- (Under guidance) apply specific treatment of the following items in the plan – data ownership and/or structure and organization, legislative restrictions, data integrity, data retrieval, data recovery, data retention, data management responsibility, data disposal and storage cost
- Create a basic database using common software packages
- Apply common database management tools specifically for use in the data maintenance processes
- Draft a template-based data maintenance plan covering, for example, the following:
  - Core structure and/or subject areas of the plan
  - Components of data architecture
#### Domain 4: Data maintenance (contd)

**Independent Practitioner**

**Knowledge Dimension**

In addition to the Novice Practitioner knowledge, be able to describe and explain:
- What a server is, the different server and data storing solutions available, and their advantages and disadvantages
- The technical (operational) process necessary to restore lost data from a back-up source
- The various debugging processes and solutions
- The components and standards of data architecture
- How to install different effective data security measures

**Skills Dimension**

In addition to the Novice Practitioner skills, be able to do the following:
- Set up a server or data-storing solutions
- Apply an authentication and security model to a database
- Maintain a database with a rigorous security model
- Resolve identified issues with the data maintenance model in place
- Use one or more data management tools/software to a basic level of proficiency
- Run routine back-ups and restore back-ups
- Activate a recovery and/or contingency plan in the event of a disaster to ensure data are not lost or data collection continues
- Identify issues with the data maintenance model in place and report/escalate identified issues to line management

**Master Practitioner**

**Knowledge Dimension**

In addition to the Independent Practitioner knowledge, be able to describe and explain:
- The factors related to each of the key focus areas: data organization, coding, storing, preserving, archiving, sharing, and data sensitivity, security, protection and integrity
- The cost components of maintaining datasets including licences and/or hidden costs and how to rationalize them
- Institutional policies on data protection, data sharing protocols and data maintenance
- The components of a comprehensive, integrated and practical disaster recovery plan that would be activated in the event of a natural or other disaster (being a formal governance document, thus requiring approval of executive management or the board – depending on the management structure of the organization)

**Skills Dimension**

In addition to the Independent Practitioner skills, be able to do the following:
- Evaluate if an existing database is consistent with standards, identify areas of concern and make the necessary changes
- Anticipate the potential of a cyber-hacking
- Draft a comprehensive cyber-attack preparedness plan
- Raise the standards of institutional data maintenance to levels of best international practices
- Make a database interoperable with other in-house or external databases as per data-sharing protocols
- Draft a comprehensive set of data protection policies
Competency Area 2
Data Processing

- Data entry
- Data cleaning
- Data validation
- Data verification
- Data transformation
<table>
<thead>
<tr>
<th>#</th>
<th>DOMAINS</th>
<th>DEFINITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data entry</td>
<td>Data entry is the mechanical process of direct entry and/or transcribing data records (often from paper-based sources) and/or audio into a data management system by means of keyboard entry or other technological processes (11).</td>
</tr>
<tr>
<td>2</td>
<td>Data cleaning</td>
<td>Data cleaning is the process of examining data to identify blemishes in datasets (for example, wrong characters, incorrect spacing and incomplete, inaccurate, incorrect, inconsistent, irrelevant or unreliable data in a dataset) and then correcting, restoring or removing the offending data prior to processing (11).</td>
</tr>
<tr>
<td>3</td>
<td>Data validation</td>
<td>Data validation consists of a series of documented data tests to ensure the validity (that is relevance, appropriateness, reliability, sourcing) and suitability of the data being reviewed (14).</td>
</tr>
<tr>
<td>4</td>
<td>Data verification</td>
<td>Data verification is the process of checking data to ensure and confirm by examination and provision of objective evidence that the data being reviewed are accurate, reliable and precise to the necessary level of detail and are consistent with data quality standards expected (that is that specified requirements have been fulfilled) (11–13).</td>
</tr>
<tr>
<td>5</td>
<td>Data transformation</td>
<td>Data transformation is the process of changing the format, structure or values of data with the purpose of making the data more clear, accurate, usable and useful (10).</td>
</tr>
</tbody>
</table>
**Domain 1: Data entry**

**KNOWLEDGE DIMENSION**

**Learner-Beginner**

- How a data system is structured, how it works and the role of data entry in the data management cycle
- The different types of data (qualitative and quantitative) and the subtypes of each one, including what is a variable, observation, datum/data point, and what are the different types of quantitative variables (numeric, logical, character, complex)
- The difference between a data-point value (content of field) and a label (description of that data field)
- Basic-level data-entry features and capabilities of Microsoft Excel and Word
- The definition of metadata and its function in the data management cycle (that is, the characteristics of data elements and their implications for data entry, for example, dates should be a standard format yyyy-mm-dd, percentages should not be above 100, names should be input as text and conform to naming protocols and standards)

**SKILLS DIMENSION**

**Learner-Beginner**

- Complete basic-level data entry into the system (Excel and Word), timely, accurately and to required standards
- Correctly identify whether data errors are systematic or random
- Record, collect and report the data points in a logbook, tally, data entry form or any other data entry tool, accurately and within expected time frames

**Novice Practitioner**

- The concepts of reporting completeness (timeliness, latency and consistency) and their application in data entry and data entry planning
- The most common challenges related to data entry on computer and paper-based platforms and for specific datasets (depending on the job requirements)

**KNOWLEDGE DIMENSION**

In addition to the Learner-Beginner knowledge, **be able to describe and explain**:

- The data entry flow from paper-based sources to digitalization and the implications for data entry, how to refer to original documentation to enter missing data

**SKILLS DIMENSION**

In addition to the Learner-Beginner skills, **be able to do the following**:

- Identify and overcome the most common challenges related to data entry on both digital and paper-based platforms and for specific datasets (depending on the job requirements)
- Complete a data entry flow from paper-based sources to digitalization
- Make point corrections to errors due to random chance and recognize when to escalate for assistance (for example, systemic errors that they are unable to fix)
- Identify whether the bug or fault is in the system or in the data, and dependent upon the finding, take appropriate action – for example, self-correct or escalate
- Recover "lost" or "corrupted" data and address system failure issues
- Access original documentation to enter missing data
- Transfer data from the computer-assisted personal interview (CAPI) data collection system to the appropriate data management system
Domain 1: Data entry (contd)

**Independent Practitioner**

**KNOWLEDGE DIMENSION**
In addition to the Novice Practitioner knowledge, **be able to describe and explain**:

- How to source comparable data from paper (logbook/tally) or digital sources
- How to create a template/empty database in which data are entered (structure of the dataset and explain how it is different to simple data entry)
- How to transfer stored data from the data collection software/source system to the appropriate data management system data warehouse
- The essence of potential specialty areas such as:
  - What matrix variables are (for network data)
  - Geographic information system (GIS) specific variables or data
  - What CAPI data collection systems are and how data are collected and stored

**SKILLS DIMENSION**
In addition to the Novice Practitioner skills, **be able to do the following**:

- Create a template/empty database in which the data are entered (structure of the dataset)
- Complete basic tasks in potential specialty areas:
  - Using matrix variables (for network data)
  - Using GIS-specific variables or data
  - Collecting and storing data using CAPI data collection systems
- Identify, report and explain problems with data entry in a manner that is standard, clear and intelligible
- Instruct junior staff and review data entry for matrices
- Transfer stored data from the data collection software/source system to the appropriate data management system data warehouse
- Fix systemic errors (for example, by mass dataset update) and use system tools to reduce their future occurrence (for example, data validation rules, applying correct data types)

**Master Practitioner**

**KNOWLEDGE DIMENSION**
In addition to the Independent Practitioner knowledge, **be able to describe and explain**:

- The different international standards available and used for specific datasets
- How to identify gaps and overlaps in data entry tools and correct them to streamline the data entry process
- Where and how appropriate support and direction should be provided to other practitioners

**SKILLS DIMENSION**
In addition to the Independent Practitioner skills, **be able to do the following**:

- Apply relevant international standards for specific and appropriate datasets
- Identify gaps and overlaps in data information and data entry tools
- Design data entry systems that respond to current and future data entry needs
- Train junior staff to prevent random and systemic errors using current best practice
## Domain 2: Data cleaning

### KNOWLEDGE DIMENSION

**Learner-Beginner**

- How to identify and remove duplicates from datasets
- The rationale for and importance of recording the data cleaning process (logging data cleaning actions) including what issues have been identified and resolved
- The meanings of "error", "implausible value" and "outlier" and the differences between them conceptually
- The role and importance of data cleaning in the data management cycle

**Novice Practitioner**

In addition to the Learner-Beginner knowledge, be able to describe and explain:

- The factors that influence reporting completeness (timeliness, latency and consistency) and how they are factored into the (data) cleaning process
- What diagnostic filters/screens are and how/when they are variously used in the data cleaning process
- The difference between an "error", "implausible value" and "outlier" and how to deal appropriately with each one
- Intermediate-to-advanced-level data cleaning functions and capabilities in Excel and Word
- The different methods used for identification of inconsistencies in data (for example, convert data types, standardized capitalization, random white spaces, inconsistent formatting, spelling errors) and how to correct them

### SKILLS DIMENSION

**Learner-Beginner**

- Highlight data using colours
- Format a "line list" and "pivot table"
- Identify and remove duplicates from datasets

**Novice Practitioner**

In addition to the Learner-Beginner skills, be able to do the following:

- Apply the various diagnostic filters/screens used in the data cleaning process
- Apply the different methods for identifying inconsistencies in data (convert data types, standardized capitalization, random white spaces, inconsistent formatting, spelling errors, etc.) and be able to correct them
- Identify data inconsistencies and identify and remove (deduplicate) datasets using different deduplication methods
- Select and use the most appropriate cleaning tools and techniques for the specific data cleaning task
- Complete data cleaning by applying the core cleaning steps/tools and using appropriate software or functions – at least District Health Information System 2 (DHIS2)/Excel, Word, if, VLOOKUP, INDEX – for the specific task
- Synthesize and report key data quality issues resolved
Domain 2: Data cleaning (contd)

**Independent Practitioner**

**KNOWLEDGE DIMENSION**
In addition to the Novice Practitioner knowledge, be able to describe and explain:
- When to pass on data changes to be approved in the data (for example, delete or replace observations and data points)
- How to source and/or create diagnostic filters/screens and streamline the data cleaning process

**SKILLS DIMENSION**
In addition to the Novice Practitioner skills, be able to do the following:
- Source and/or create diagnostic filters/screens and streamline the data cleaning process
- Source data cleaning rules if documents are missing
- Identify and authenticate potential data errors and data outliers
- Draft a comprehensive data cleaning report

**Master Practitioner**

**KNOWLEDGE DIMENSION**
In addition to the Independent Practitioner knowledge, be able to describe and explain:
- How to approve changes to the data
- The features and characteristics of sensitive data
- How to anonymize datasets
- How to identify systematic and random errors in data quality and streamline the data cleaning process

**SKILLS DIMENSION**
In addition to the Independent Practitioner skills, be able to do the following:
- Create data diagnostic filters and quality screens
- Approve changes to data (for example, delete or replace observations and data points)
- Identify sensitive data and anonymize datasets
- Identify systematic and random errors in data quality and streamline the data cleaning process
Domain 3: Data validation

**KNOWLEDGE DIMENSION**

**Learner-Beginner**

*Be able to describe and explain:*
- What data validation is and its role and importance in the data entry process
- The rules for capturing data consistently and why these rules are important
- The key characteristics used to define data quality (that is validity, accuracy, completeness, consistency, uniformity)

**SKILLS DIMENSION**

*Be able to do the following:*
- Complete basic data validation using standard rules/guidelines in place for common attributes (for example, telephone number, names, identification) and the key data quality characteristics (for example, validity, accuracy, completeness, consistency, uniformity)
- Document the data validation process using a standard reporting format
- Implement preliminary data validation checks as directed by senior staff

**Novice Practitioner**

*In addition to the Learner-Beginner knowledge, be able to describe and explain:*
- The diagnostic tests/screens available for data validation, their specific purposes and how to use them for common attributes (for example, telephone number, names, identification)
- The purpose and rationale for recording data validation actions and how to document the validation process in a standard reporting format
- The data validation rules and diagnostic tests/screens for any type of data, their specific purposes and how to use them, comparing data points systematically and rigorously
- How to refer to original data sources to correct data quality issues

*SKILLS DIMENSION*

*In addition to the Learner-Beginner skills, be able to do the following:*
- Use diagnostic tests/screens for any type of data, comparing data points systematically and rigorously
- Revert to original data sources to correct data quality issues
- Set limits for possible numeric values for a given field and enter values that fit within a list or range of acceptable values
- Complete a data quality validation check based on the eight characteristics of data validation (originality, attribution, accuracy, consistency, legibility, contemporaneousness, endurance, completeness)
- Review source of data points (for example, paper-based documentation) to solve potential data quality issues based on the data validation findings
- Compare data points systematically and rigorously
<table>
<thead>
<tr>
<th>Level</th>
<th>KNOWLEDGE DIMENSION</th>
<th>SKILLS DIMENSION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Independent Practitioner</strong></td>
<td>In addition to the Novice Practitioner knowledge, be able to describe and explain:</td>
<td>In addition to the Novice Practitioner skills, be able to do the following:</td>
</tr>
<tr>
<td></td>
<td>• When to approve changes in the data according to the data quality issues identified</td>
<td>• Run data validation checks for any type of data</td>
</tr>
<tr>
<td></td>
<td>• How to identify the difference between systematic and random errors in data quality</td>
<td>• Recognize when changes are needed and make decisions about data changes based on the data quality issues identified and on the validation findings (post-validation checks actions)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Identify and fix systematic and random errors in data quality</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Document the post-validation checks processes on a standard reporting format</td>
</tr>
<tr>
<td><strong>Master Practitioner</strong></td>
<td>In addition to the Independent Practitioner knowledge, be able to describe and explain:</td>
<td>In addition to the Independent Practitioner skills, be able to do the following:</td>
</tr>
<tr>
<td></td>
<td>• The process of automating data validation tasks and streamlining the data validation process</td>
<td>• Automate data validation tasks and streamline the data validation process</td>
</tr>
<tr>
<td></td>
<td>• The steps in designing a data validation process that recognizes health, ethical and regulatory concerns</td>
<td>• Create and run a data validation process that recognizes health, ethical and regulatory concerns, using appropriate data management software covering the following phases: planning, data validation, entry, (database) lock</td>
</tr>
<tr>
<td></td>
<td>• The process of designing and running tests to identify systematic and random data quality errors</td>
<td>• Design and run tests to identify systematic and random data quality errors</td>
</tr>
<tr>
<td></td>
<td>• When additional validation processes are required, and the role of data quality statements</td>
<td>• Create relevant and appropriate data validation tasks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Design and run checks to determine if changes are necessary to produce the analysis datasets either through the automated facilities of a data dictionary, or by the inclusion of explicit programme validation logic</td>
</tr>
</tbody>
</table>

**Domain 3: Data validation (contd)**

**AREA 2: Data Processing**

**COMPETENCY AREA 2: DATA PROCESSING**
# Domain 4: Data verification

## KNOWLEDGE DIMENSION

**Learner-Beginner**

**Be able to describe and explain:**
- What data verification is and its role and importance in the data entry cycle
- The purpose and rationale for documenting the verification process
- The key criteria used in the data-verification process
- How to document the verification process in a standard reporting format

## SKILLS DIMENSION

**Be able to do the following:**
- Complete verification tasks using each of the four verification methods (inspection, demonstration, test, analysis) under supervision
- Document the verification process on a standard reporting format

## Novice Practitioner

**KNOWLEDGE DIMENSION**

In addition to the Learner-Beginner knowledge, **be able to describe and explain:**
- How to migrate, merge and append data to be able to identify potential data quality issues for data verification

**SKILLS DIMENSION**

In addition to the Learner-Beginner skills, **be able to do the following:**
- Migrate, merge and append data to be able to identify potential data quality issues for data verification
- Document the data-verification processes on a standard reporting format

## Independent Practitioner

**KNOWLEDGE DIMENSION**

In addition to the Novice Practitioner knowledge, **be able to describe and explain:**
- What steps to take to correct potential data quality issues arising during data verification

**SKILLS DIMENSION**

In addition to the Learner-Beginner skills, **be able to do the following:**
- Correct data quality issues arising during data verification
- Design and complete both manual and/or automated data sampling and checking exercises to verify the information in the destination system matches the source system
- Use artificial intelligence and/or machine learning for data verification
- Document the post-verification checks processes on a standard reporting format
Domain 4: Data verification (contd)

**KNOWLEDGE DIMENSION**

In addition to the Independent Practitioner knowledge, **be able to describe and explain**:

- How to create or adapt new best practices using current or new tools for data verification. For example, use artificial intelligence/machine learning
- How to automate data verification tasks to streamline the data verification process

**SKILLS DIMENSION**

In addition to the Independent Practitioner skills, **be able to do the following**:

- Automate data verification tasks to streamline the data verification process
- Quality control the verification activities in a data migration/merge exercise to ensure the correct transfer of data from source location
- Intervene in the recurring quality assurance process with appropriate verification intervention methodology
- Design data verification tools and/or code
- Create, adopt or adapt best practices using current or new tools for data verification
# Domain 5: Data transformation

## KNOWLEDGE DIMENSION

**Learner-Beginner**

- Be able to describe and explain:
  - The different types of data transformation and when and how they are applied in the data transformation process.
  - The objective of the intended data transformation, the process required, the utility of the data elements in the dataset and how they should be combined.
  - The different data formats, their extraction for data generation and transformation (for example, csv, json) and how to change the data format.
  - The difference between a "pivot" and a "line-list" file and their implication for data transformation.
  - What the required data elements to process the indicator are (for example, numerator/denominator), whether the data elements are available and usable in the dataset and how they should be combined.
  - The purpose, benefits and challenges of data transformation.

## SKILLS DIMENSION

**Learner-Beginner**

- Be able to do the following:
  - Complete basic data transformation tasks using appropriate software – for example, Microsoft solutions, R, Stata, Statistical Package for the Social Sciences (SPSS), statistical analysis system (SAS), Python – under supervision.
  - Produce the appropriate dataset for data transformation tasks.
  - Provide support for the creation of metadata files.

## KNOWLEDGE DIMENSION

**Novice Practitioner**

- In addition to the Learner-Beginner knowledge, **be able to describe and explain**:
  - How to import and export datasets in various formats.
  - The purpose and rationale of documenting the data transformation process used.
  - How to document the transformation process in a standard reporting format.
  - The meanings of “unique identifier”, “key variable”, “data merging” and “appending/binding” and their purpose.
  - The different shapes of datasets (long, wide) and how to convert tabular data to row data.
  - How to code basic data transformations.
  - The difference between “data” and “metadata”.
  - The process of creating metadata files based on certain data transformations (for example, how an indicator was calculated).

## SKILLS DIMENSION

**Novice Practitioner**

- In addition to the Learner-Beginner skills, **be able to do the following**:
  - Import and export datasets in appropriate formats.
  - Complete more complex data transformation tasks, set up a clean data transformation process (for example, store working files correctly, separate raw and processed data) using Excel, with minimal supervision, and transfer stored data from the data collection software/source system to the appropriate data management system data warehouse (for example, extract data from DHIS2 or other platforms).
  - Document the data transformation process in a standard reporting format.
  - Alter data formats and reshape datasets (tabular to row data and vice versa) correctly.
  - Code basic data transformations.
  - Create metadata files based on certain data transformations (for example, how an indicator was calculated).
  - Merge or append/bind datasets based on unique identifiers.
**Domain 5: Data transformation (contd)**

### KNOWLEDGE DIMENSION

**Independent Practitioner**

In addition to the Novice Practitioner knowledge, **be able to describe and explain**:

- The difference between the extract, transform and load (ETL) and the extract, load and transform (ELT) models of data transformation and the factors that determine which transformation model is best suited to which circumstance
- The master lists in place for essential health components (for example, village, hospital) and how to use these lists in a data transformation process
- What a database management system (DBMS) is and associated data architecture, including relational and non-relational databases
- How to code advanced data transformations
- The core functions of structured query language (SQL) and other query languages
- How to recognize sensitive/confidential data and how to anonymize a dataset

**Skills Dimension**

In addition to the Novice Practitioner skills, **be able to do the following**:

- Define the objective of the intended data transformation, identify the required data elements to process the indicator (for example, numerator/denominator), identify whether the data elements are available and usable in the dataset and combine them effectively
- Use both ETL and ELT models of data transformation, selecting the most appropriate transformation model
- Create and use the master lists in place for essential health components (villages, hospital) in a data transformation process
- Code advanced data transformations
- Use a DBMS and associated data architecture, including relational and non-relational databases
- Use the core functions of SQL and other query languages
- Recognize sensitive/confidential data and anonymize a dataset accordingly
- Complete more advanced data transformation tasks such as "data type conversion" and "hierarchical data flattening" and complete larger-scale data-transformation tasks using scripting languages such as Python or R

### Master Practitioner

**Knowledge Dimension**

In addition to the Independent Practitioner knowledge, **be able to describe and explain**:

- How to create and/or adopt new best practices using current or new tools for data transformation (for example, design R code to automatically prepare system output data structure for further analysis)
- How to automate data transformation tasks to streamline data processing activities

**Skills Dimension**

In addition to the Independent Practitioner skills, **be able to do the following**:

- Automate data transformation tasks to streamline data processing activities
- Train junior staff on the best practice on a given software (usually Excel or SPSS for basic transformations, and Python or R for bulk transformation tasks)
- Create and/or adopt best practices using current or new tools for data transformation (for example, design R code to automatically prepare system output data structure for further analysis)
Competency Area 3
Data Analysis

- Descriptive analysis
- Diagnostic analysis
- Predictive analysis
- Prescriptive analysis
## Competency domains

<table>
<thead>
<tr>
<th>#</th>
<th>Domains</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Descriptive analysis</td>
<td>Descriptive analysis is the process of using current and historical data to identify trends and relationships to provide meaningful information on current or recent events by converting raw data into a form and format that attempts to answer who, what, where and when questions (15,16,18). Descriptive analysis is the cornerstone of data analysis and the starting point for all data insights. It may provide the first ideas about the subject of the analysis, for example, counts of the age or sex of the cohort under study.</td>
</tr>
<tr>
<td>2</td>
<td>Diagnostic analysis</td>
<td>Diagnostic analysis deals with causality and thus generates and tests hypotheses that provide crucial information about why a trend or relationship occurred. It is concerned with the search for identifying and understanding causation versus correlation, hypothesis testing, proving assumptions, etc., such that the analysis delivers insights that are both non-obvious and value-added to the analysis process (15,17). An example could be hypothesizing and verifying whether the age or gender of the cohort under study is truly correlated with an outcome of interest.</td>
</tr>
<tr>
<td>3</td>
<td>Predictive analysis</td>
<td>Predictive analysis is the use of data and techniques to predict (or identify) the likelihood of future outcomes. Predictive analysis draws on past and current data and often uses comprehensive and sophisticated methods (for example, statistical algorithms and machine learning techniques) to predict the likelihood of future outcomes (15,19). The goal is to provide a prediction of what will happen in the future. An example might be the use of age and sex as input variables to a model that calculates the likelihood of purchase behaviour to increase accuracy.</td>
</tr>
<tr>
<td>4</td>
<td>Prescriptive analysis</td>
<td>Prescriptive analysis mainly focuses on the process of using data to determine an optimal course of action. It is the most advanced, sophisticated and challenging analysis domain. Prescriptive analysis combines the outcomes from the other three analysis domains and factors in information about available resources, past and current performance, possible situations and wider environmental factors, and applies them to the process of decision-making, to assist decision-makers in determining the optimum solution from a variety of available options (15,20). This is a complex and highly specialized type of analysis, the expertise for which is most commonly found in companies and organizations that provide specialist, third-party commercial services in this domain. An example might be the use of age and sex of all users of a social media site as input variables to an algorithm that maximizes watch-time and engagement.</td>
</tr>
</tbody>
</table>
Domain 1: Descriptive analysis

**KNOWLEDGE DIMENSION**

**Learner-Beginner**

**Be able to describe and explain:**

- Basic statistical concepts of central tendency (for example, average, median, mode), dispersion (for example, standard deviation) and disease frequency (for example, per cent, rate, ratio, prevalence, incidence) for the description of quantitative data

- The most appropriate choice of graph/figure (for example, bar chart, pie chart, line graph, map) for each data element or indicator, including the reasoning behind the selection

**SKILLS DIMENSION**

**Be able to do the following:**

- Identify and extract key messages and findings from a given descriptive analysis exercise
- Complete a basic central tendency (statistical) analysis exercise (for example, calculate average, median, mode), dispersion (for example, standard deviations), and disease frequency (for example, per cent, rate, ratio, prevalence, incidence, etc.) for description of quantitative data

**Novice Practitioner**

**In addition to the Learner-Beginner knowledge, be able to describe and explain:**

- Basic interpretations of data visualizations with one or two layers of data information (for example, rendering an indicator as three-line graphs – representing total, male and female populations – over time can highlight patterns between sexes that an aggregated graph cannot)

- Potential data quality issue, such as the concept of outliers, data quality and statistical inconsistencies, and how they affect the results of descriptive analysis

**SKILLS DIMENSION**

**In addition to the Learner-Beginner skills, be able to do the following:**

- Create various types of statistical tables and select the situations where they may be most usefully applied (for example, 2x2 table for false positivity rates)

- Select and apply the most appropriate choice of graph/figure (for example, bar chart, pie chart, line graph, map) for each data element or indicator, for a given exercise and explain the reasoning behind selection

- Complete a basic descriptive analysis task (typically collation, summarization, aggregation, structuring and organization of data using appropriate techniques, for example, summarizing using sum of counts, and weighted average for rates, drawing 2x2 tables to describe false positivity rate of screening method vs a gold standard method)

- Visualize and present findings in a logical, coherent and meaningful way, appropriate to a given case

- Select and apply the most appropriate data elements and indicators, especially relevant key performance indicators, in a given workstation, and use the correct statistical techniques to calculate them (for example, prevalence of current tobacco use should be age-standardized, the denominator of vaccination coverage should be target population and not always the entire population in country, etc.)

- The different types of statistical tables and the situations where they may be useful (for example, 2x2 table for false positivity rates)

- Complete a basic data visualization task with one or two layers of data information (for example, three-line graphs to show trend of indicator over time between total, male and female populations using Excel)

- Recognize and detect data quality issues, such as outliers, data quality and statistical inconsistencies, and identify how they affect the results of a descriptive analysis process
Domain 1: Descriptive analysis (contd)

**KNOWLEDGE DIMENSION**

**Independent Practitioner**

In addition to the Novice Practitioner knowledge, **be able to describe and explain**:

- The software typically used at their workstation to perform descriptive analysis and how they are used (for example, DHIS2 dashboards for monitoring immunization coverage and dropout rate, Excel)
- Basic qualitative methods and the situations where they are used (for example, observations, textual or visual analysis of books or videos, key informant interviews and focus group discussions)
- The results of descriptive analysis correctly and accurately using the proper units, terminology, disclaimers (for example, measles coverage for the district is 90% based on available data from 10/15 health centres)

**SKILLS DIMENSION**

In addition to the Novice Practitioner skills, **be able to do the following**:

- Perform descriptive analysis using the software typically used at their workstation and demonstrate how they are used (for example, DHIS2 dashboards for monitoring immunization coverage and dropout rate, Excel, etc.)
- Identify statistical inconsistencies before, during and after analysis and perform the correct follow-up action to address them
- Assist in the creation, selection and use of analysis outputs in data-dashboards, high-level reports and other data products
- Describe the results of a descriptive analysis exercise, accurately, using the proper units, terminology, disclaimers (for example, measles coverage for the district is 90% based on available data from 10/15 health centres)

**Master Practitioner**

In addition to the Independent Practitioner knowledge, **be able to describe and explain**:

- Best practices for summarizing and visualizing data and indicators relevant to the workstation (for example, component bar-graphs should be used for parts of a whole, maps [animated for years] for trends over time)
- The newest features of tools/software for developing and enhancing data summarization and visualization practices
- The relevant analytical features of other statistical software such as R, Stata, SPSS, SAS, Python, etc. and their current or potential application to a workstation’s activities (for example, R script for automatically downloading and cleaning of data from DHIS2, Python for rendering online visualization for dissemination)

**SKILLS DIMENSION**

In addition to the Independent Practitioner skills, **be able to do the following**:

- Effectively use relevant analytic features of other statistical software such as R, Stata, SPSS, SAS, Python, etc. recognizing their current or potential application to a workstation’s activities (for example, R script for automatically downloading and cleaning data from DHIS2, Python for rendering online visualization for dissemination, etc.)
- Conceptualize and design data-dashboards, high-level reports and other data products appropriate for a given use case
- Apply best practices for summarizing and visualizing data and indicators relevant to the workstation (for example, component bar-graphs should be used for parts of a whole, maps [animated for years] for trends over time)
- Experiment with the newest features of tools and software to develop and enhance data summarization and visualization practices
Domain 2: Diagnostic analysis

**Learner-Beginner**

**KNOWLEDGE DIMENSION**

Be able to describe and explain:

- The fundamentals and basic methods of diagnostic analysis

**SKILLS DIMENSION**

- Not a required skill set at this level

**Novice Practitioner**

**KNOWLEDGE DIMENSION**

In addition to the Learner-Beginner knowledge, be able to describe and explain:

- The most suitable methods of diagnostic analysis that should be applied based on different data types, research design and analysis purpose (methods include statistical methods or other broader methods in both qualitative and quantitative areas, data visualizations, etc.)
- Basic statistical theories (such as P value, confidence interval)

**SKILLS DIMENSION**

Be able to do the following:

- Select the most suitable methods of diagnostic analysis, which should be applied based on data types, research design and analysis purpose. The methods include statistical methods or other broader methods in both qualitative and quantitative areas, data visualizations, etc.

**Independent Practitioner**

**KNOWLEDGE DIMENSION**

In addition to the Novice Practitioner knowledge, be able to describe and explain:

- How to identify true differences (for example, between samples and targeted population) or test hypotheses based on a univariate (for example, t-test, Chi-square test, non-parametric test)
- How and under what circumstances advanced-level data visualization techniques are used
- How advanced software is used to generate figures/graphs with multi-layer results (for example, Tableau, R, GIS)

**SKILLS DIMENSION**

In addition to the Novice Practitioner skills, be able to do the following:

- Identify true differences (for example, between samples and targeted population) or test hypothesis based on a univariate (such as t test, Chi-square test, non-parametric test)
- Make correct inferences from current data to broader more general conditions (21)
- Apply the chosen statistical methods to a specific task and context.
Area 3: Data Analysis

In addition to the Independent Practitioner knowledge, be able to describe and explain:

- How to identify potential impact and risk factors, causal and non-causal relationships (probability, likelihood) based on multi-variates and multi-levels by using advanced statistical methods (for example, correlation, regression, multi-way analysis of variance [ANOVA], survival analysis)
- The various stages, steps, procedures and challenges for a given qualitative analysis task
- How to make correct inferences from advanced statistical results and how to apply these inferences to broader more general scenarios

In addition to the Independent Practitioner skills, be able to do the following:

- Identify potential impact and risk factors, causal and non-causal relationships (probability, likelihood) based on multi-variates and multi-levels using advanced statistical methods (for example, correlation, regression, multi-way ANOVA, survival analysis)
- Select and apply advanced qualitative methods (grounded theory, ethnography and participant observation, visual methods, classical content analysis, experience sampling techniques, rhetorical analysis)
- Make correct inferences from advanced statistical results and apply them to broader, more general scenarios
- Complete a qualitative data analysis exercise demonstrating the various stages, steps, procedures and challenges for current analysis
- Select and apply the latest methods and newest features of tools/software for better diagnostic outcomes for same/similar areas
## Domain 3: Predictive analysis

### Learner-Beginner

#### KNOWLEDGE DIMENSION

**Be able to describe and explain:**

- The purpose, objective and wider context of a planned workstation analysis project and the meaning of the expected optimizations (for example, immunization unit aims to maximize immunization coverages and minimize dropout rates and vaccine vial wastage).

#### SKILLS DIMENSION

- Not a required skill set at this level

### Novice Practitioner

#### KNOWLEDGE DIMENSION

In addition to the Learner-Beginner knowledge, **be able to describe and explain:**

- The concept of simple prediction and rudimentary methods for performing them (for example, extrapolating future points using arithmetic or geometric average change per year).
- The ideal qualities of a dataset (for example, selecting correct parameters/variables) for use in statistical modelling and the process of preparing these using data in a given workstation.
- The general statistical principles of decision trees and regression* and their potential for application in predictive analysis.

#### SKILLS DIMENSION

- Not a required skill set at this level

---

* "Decision trees" are models for identifying the variable or set of variables that will split a dataset into the most different groups. The selection and number of variables, as well as the criteria applied for splitting used in the model, should be purposeful (for example, too many low-value variables risk compromising the predictive power of the model in a real-world setting) (20). "Regression" is a set of models for estimating the relationships between variables, has different types (simple and multiple, linear and logistic, etc.) each with their own set of assumptions (main ones being normal distribution and multi-collinearity for linear and logistic regression respectively), and appropriate interpretations (20).
In addition to the Novice Practitioner knowledge, **be able to describe and explain**:

- What longitudinal analysis is and its capability to identify data patterns and trends
- How anomalies and/or contradictions can arise in data and how to recognize them
- The process of extrapolating to predict future outcomes while adequately communicating assumptions and limitations in the chosen analytic method
- The meanings of "linear regression", "inferential analysis", "classification" and "data segmentation"
- The general statistical principles of predictive analysis exercises such as decision trees and regression
- The relevant datasets within a given workstation and their potential to be used for relevant predictive analyses (for example, extract regularly collected facility-based immunization database to forecast demand and avoid stockouts or wastage)
- The value-added of different datasets to a given analysis work (for example, combining multidrug-resistant tuberculosis (MDR-TB) case dataset with health insurance database to add or verify information on co-morbidities)
- The basic concepts and rudimentary methods of simple predictive analysis (for example, extrapolating future points using arithmetic or geometric average change per year)
- The action points needed to design a study that will feature regression or decision trees (for example, including variables relevant to logistic regression in data collection activities to allow analysis in the future)
- The relevant datasets within a given workstation and their potential to be used for relevant predictive analyses (for example, extract regularly collected facility-based immunization database to forecast demand and avoid stockouts or wastage)
- The process of extrapolating to predict future outcomes while adequately communicating assumptions and limitations in the chosen analytic method
- The process of extrapolating to predict future outcomes while adequately communicating assumptions and limitations in the chosen analytic method
- The action points needed to design a study that will feature regression or decision trees (for example, including variables relevant to logistic regression in data collection activities to allow analysis in the future)
- The general statistical principles of predictive analysis exercises such as decision trees and regression
- The relevant datasets within a given workstation and their potential to be used for relevant predictive analyses (for example, extract regularly collected facility-based immunization database to forecast demand and avoid stockouts or wastage)
- The value-added of different datasets to a given analysis work (for example, combining multidrug-resistant tuberculosis (MDR-TB) case dataset with health insurance database to add or verify information on co-morbidities)
- The necessary software and corresponding commands – for example, rpart.plot, integrated nested Laplace approximation (INLA) in R; pandas, numpy and matplotlib in Python – to efficiently run regression, decision trees and other analyses on relevant variables

**SKILLS DIMENSION**

**Be able to do the following:**

- Complete a predictive analysis exercise demonstrating the application of basic concepts and rudimentary methods of simple predictive analysis (for example, extrapolating future points using arithmetic or geometric average change per year)
- Select the correct parameters for a dataset for use in statistical modelling and complete the process of preparing these using data
- Carry out basic linear or logistic regression, inferential analysis, classification and data segmentation tasks
- Complete a predictive analysis exercise demonstrating the application of general statistical principles of decision trees and regression
- Select appropriate datasets and correctly apply them to a relevant predictive analysis task within a given workstation (for example, extract regularly collected facility-based immunization database to forecast demand and avoid stockouts or wastage)
- Design a study that will feature regression or decision trees (for example, incorporate child nutrition modules in periodic mother and child surveys to be able to add child weight classification as variables in logistic regression)
- Analyse past data demonstrating the capability to identify patterns and trends, while detecting anomalies and/or contradictions and extrapolate from the above to predict future outcomes while adequately communicating assumptions and limitations in the chosen analytic method
- Select and appropriately combine a variety of datasets to a specific analysis task (for example, combining MDR-TB case dataset with health insurance database to add or verify information on co-morbidities)
- Run accurate regression, decision trees and other analyses on relevant variables using the necessary software and corresponding commands (for example, rpart.plot, INLA in R; pandas, numpy and matplotlib in Python)
Domain 3: Predictive analysis

**KNOWLEDGE DIMENSION**

In addition to the Independent Practitioner knowledge, **be able to describe and explain**:

- The most effective predictive analysis solutions in health context issues and viable options that can be proposed to senior management and other stakeholders
- The technical differences between variations of advanced predictive analysis techniques for a given analysis task
- The benefits, potential disadvantages, concerns and limitations of these advanced data science techniques (for example, machine learning, heuristics, neural network) for forecasting work in a given workstation
- The newest methods and features of tools/software and their potential for improving current predictive analysis procedures or subject areas
- The advantages and disadvantages of used methods and modelling techniques to guide independent and advanced practitioners to select the most appropriate one and effect proper adjustment (for example, Bayesian analysis, Poisson regression, survival analysis, time series data mining)
- The process for transferring the conceptual design of an advanced statistical model into a set of practical usable instructions to guide real-world decisions to design a notification app to warn of high-transmission area threats (for example, model for probability of infection by a particular disease, given current location and other socio-demographic factors)

**SKILLS DIMENSION**

In addition to the Independent Practitioner skills, **be able to do the following**:

- Select the most effective predictive analysis solutions in specific health context issues and proactively propose viable options to senior management and other stakeholders
- Select and apply the most advanced (operationally appropriate) data science techniques (for example, machine learning, heuristics, neural network) for forecasting work in a given workstation
- Select and apply the most appropriate statistical modelling techniques in a given situation in order to understand the underlying nature of the data being analysed
- Select and apply the latest methods and newest features of tools/software for better predictive diagnostic outcomes for same/similar areas
- Select and apply the most appropriate advanced predictive analysis techniques for a given analysis work
- Select and competently apply in appropriate contexts the less commonly used methods and modelling techniques (for example, Bayesian analysis, Poisson regression, survival analysis, time series data mining) to guide independent and advanced practitioners to select the most appropriate one and effect proper adjustments
- Draft a practical users guide that translates the conceptual design of an advanced statistical model into a set of practical, usable instructions to guide real-world decisions for the design of a notification app to warn of high-transmission area threats (for example, model for probability of infection by a particular disease, given current location and other socio-demographic factors)

---

† "Statistical modelling" is a method of mathematically approximating the world. A "statistical model" is the use of statistics to build a representation of the data and then conduct analysis to infer any relationships between variables or discover insights (23). "Machine learning" is the use of mathematical and or statistical models to obtain a general understanding of the data to make predictions (24).
Domain 3:
Predictive Analysis

COMPETENCY AREA 3: DATA ANALYSIS
Domain 4: Prescriptive Analysis

**Learner-Beginner**

**KNOWLEDGE DIMENSION**

Be able to describe and explain:

- The definition of prescriptive analysis especially in contrast to the other domains of data analysis
- The contexts where prescriptive analysis is used

**SKILLS DIMENSION**

- Not a required skill set at this level

---

**Novice Practitioner**

**KNOWLEDGE DIMENSION**

In addition to the Learner-Beginner knowledge, be able to describe and explain:

- The potential applications for data analysis (for example, useful for detecting fraud, disease, quality, safety issues patterns in real time and prescribe actions)

**SKILLS DIMENSION**

- Not a required skill set at this level

---

**Independent Practitioner**

**KNOWLEDGE DIMENSION**

In addition to the Novice Practitioner knowledge, be able to describe and explain:

- The basic concepts underpinning prescriptive analysis

**SKILLS DIMENSION**

- Not a required skill set at this level

---

**Master Practitioner**

**KNOWLEDGE DIMENSION**

In addition to the Independent Practitioner knowledge, be able to describe and explain:

- How to collaborate with third-party/independent companies to conduct prescriptive analysis by correctly transferring health-related needs to data language and provide health-related inputs, guidance and concerns
- How to collect and monitor the ongoing analysis related with health areas
- How to explore potential needs for prescriptive analysis in own health areas

**SKILLS DIMENSION**

Be able to do the following:

- Collaborate effectively with third-party/independent companies to conduct prescriptive analysis by correctly transferring health-related needs to data language, able to provide health-related inputs, guidance and concerns
- Collect and monitor the ongoing analysis related with health areas
- Explore and examine potential prescriptive analysis needs for own health areas
AREA 3: Data Analysis
Competency Area 4
Data Usage

- Data interpretation and conclusion
- Data integration
- Data presentation and reporting
- Transfer to action-oriented recommendations
## Competency domains

<table>
<thead>
<tr>
<th>#</th>
<th>DOMAINS</th>
<th>DEFINITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data interpretation and conclusion</td>
<td>Data interpretation is the process of reviewing results of data analysis, making inferences, assigning meaning to the findings and producing actionable insights from which accurate and appropriate conclusions can be drawn (25). Conclusion is the process of summarizing key information points and arriving at a final judgement (on an issue) based on reasoning from facts, logic and evidence of the data provided (26,27).</td>
</tr>
<tr>
<td>2</td>
<td>Data integration</td>
<td>Data integration is the process of combining data from several disparate, heterogeneous sources into a coherent data framework to retain and support a consolidated perspective of the information gathered and obtain a rounded quantitative and/or qualitative impression of the overall effect of a particular intervention (or variable) on a defined outcome with the objective of deriving actionable insights (9). Data integration can consolidate all kinds of data – structured, unstructured, batch and streaming – to do everything from basic querying of inventory databases to complex predictive analytics.</td>
</tr>
<tr>
<td>3</td>
<td>Data presentation and reporting</td>
<td>Data presentation and reporting is the process of collecting unprocessed data from different sources and, by use of both narrative and graphical tools, techniques and formats, converting that data into meaningful information that provides valuable insights and enables informed decision-making (12).</td>
</tr>
<tr>
<td>4</td>
<td>Transfer to action-oriented recommendations</td>
<td>Transfer to action-oriented recommendations is the process of transferring evidence-derived knowledge into ethically sound, action-oriented recommendations (28). The focus is on the characteristics of the recommendations, which should give direction to the subsequent actions. Specific, measurable, achievable, relevant and time-bound (SMART) criteria can be used to estimate the quality of action-oriented recommendations.</td>
</tr>
</tbody>
</table>
# Domain 1: Data interpretation and conclusion

## Learner-Beginner

**KNOWLEDGE DIMENSION**

*Be able to describe and explain:*

- The various basic data types, terms, methods used and analysed in the results and how each method is applied in practice
- Why and how the coding (that is SAS) process (from raw data to initial results generated) is used

**SKILLS DIMENSION**

*Be able to do the following:*

- Identify the basic data types, terms, methods used and analysed in the results
- Apply each of the above (types, terms, methods) in practical contexts

## Novice Practitioner

**KNOWLEDGE DIMENSION**

In addition to the Learner-Beginner knowledge, *be able to describe and explain:*

- The meaning and application of key terms used in the process of forming conclusions (for example, inductive and deductive reasoning, premise, hypothesis, supposition, inference)

**SKILLS DIMENSION**

In addition to the Learner-Beginner skills, *be able to do the following:*

- Interpret and draw basic conclusions based on the data results including data points, text, graphs, charts and maps

## Independent Practitioner

**KNOWLEDGE DIMENSION**

In addition to the Novice Practitioner knowledge, *be able to describe and explain:*

- Advanced data types, terms, methods used and analysed in the results, how each method is applied in practice, and likely implications and consequences
- How to use data results including data points, text, graphs, charts and maps to draw accurate conclusions

**SKILLS DIMENSION**

In addition to the Novice Practitioner skills, *be able to do the following:*

- Use data results including data points, text, graphs, charts, and maps to draw accurate conclusions
- Communicate the results as applicable to different audiences
- Apply advanced data types, terms, methods used and analysed in the results, and demonstrate how each method is applied in practice
KNOWLEDGE DIMENSION

In addition to the Independent Practitioner knowledge, be able to describe and explain:

- The process, tools and techniques involved in “back-casting” the data resources, collection and analysis methods
- The limitations and/or restrictions of the selected process and predict potential data-gaps
- The different data resources, collection methods and distinguishing characteristics and predict potential limitations
- The key information/conclusion requirements for different types of audiences: higher-level managers for decision-making, public audiences, technical experts, field workers or lower-level implementation staff
- From an examination of data produced, identify the specific research and analysis processes used to deliver the data, and from that, explain potential weaknesses in the selected processes that could possibly result in potentially flawed interpretations and/or conclusions
- Links with demographically similar areas that are conducting similar studies to draw a complete picture of the conclusion
- The actual and potential information gaps that will require further data collection in the next step

SKILLS DIMENSION

In addition to the Independent Practitioner skills, be able to do the following:

- Examine data produced, identify the specific research and analysis processes used to deliver the data and, from that, identify potential weaknesses in the selected processes that could possibly result in potentially flawed interpretations and/or conclusions
- Demonstrate the ability to apply the process, tools and techniques involved in data interpretation the data resources, collection, and complex analysis methods
- Identify the processes used (to deliver the data) and, from that, identify actual weaknesses in the processes that could possibly result in flawed interpretations and conclusions
- Identify potential data gaps that need further collection in future to draw further conclusions
- Identify the processes used (to deliver the data) and from that identify actual weaknesses in the processes that could possibly result in flawed interpretations and conclusions
- Compare with demographically similar health areas to extract potential lessons and identify potential information gaps that will require further collection in the next steps
- Generate and communicate key conclusions as relevant for different audience types – stakeholders: higher-level managers for decision making, public audiences, technical experts, field workers or implementation staff at the lower level
Domain 2: Data integration

**Learner-Beginner**

**KNOWLEDGE DIMENSION**

Be able to describe and explain:
- The core concept and principles of data integration

**SKILLS DIMENSION**

Be able to do the following:
- Apply the core concepts and basic principles in a basic data integration task

---

**Novice Practitioner**

**KNOWLEDGE DIMENSION**

In addition to the Learner-Beginner knowledge, be able to describe and explain:
- How data integration processes are applied in different contexts

**SKILLS DIMENSION**

In addition to the Learner-Beginner skills, be able to do the following:
- Apply basic data integration processes in various integration contexts

---

**Independent Practitioner**

**KNOWLEDGE DIMENSION**

In addition to the Novice Practitioner knowledge, be able to describe and explain:
- The importance of the data integration process in crafting the full story
- The other data requirements needed to integrate data and draw clear conclusions
- How conclusions drawn may impact routine work processes

**SKILLS DIMENSION**

In addition to the Novice Practitioner skills, be able to do the following:
- Integrate current findings to inform actions with the aim to improve the effectiveness of current routine work practices
Domain 2: Data integration

**KNOWLEDGE DIMENSION**

In addition to the Independent Practitioner knowledge, be able to describe and explain:

- What data are available from other sources and the benefit of accessing those data
- Data integration methodologies (that is data linkage and associated challenges)
- Potential limitations of not adopting integrative methods (that is data linkage)
- How to accurately draw fully informed conclusions

**SKILLS DIMENSION**

In addition to the Independent Practitioner skills, be able to do the following:

- Identify the different data requirements needed to integrate data and draw clear accurate conclusions
- From results/data produced, examine the quality of integration applied and from that to identify potential weaknesses that could result in flawed interpretations and conclusions
- Complete a systematic review of peer-reviewed and grey literature and present the aggregated and summarized findings from that review
- Compare multi-source data to draw accurate and fully informed conclusions
## Domain 3: Data presentation and reporting

### Learner-Beginner

**KNOWLEDGE DIMENSION**

**Be able to describe and explain:**

- The core concepts of data presentation and reporting (that is the key factors that determine the content and structure of a presentation/report – purpose, objectives, context, audience needs and expectations), including which format, tools and techniques are most appropriate for both standard and non-standard presentations and reports
- The appropriate criteria to be used to determine the contents of a standard presentation/report
- The core principles of integrity in data reporting and how to apply them in practice

**SKILLS DIMENSION**

**Be able to do the following:**

- Identify whether the core concepts of data presentation and reporting are evident and presented in a standard report/presentation
- Identify whether core integrity principles have been applied in a particular report/presentation
- Identify whether or not the appropriate content selection criteria have been applied in a particular case

### Novice Practitioner

**KNOWLEDGE DIMENSION**

In addition to the Learner-Beginner knowledge, **be able to describe and explain:**

- How the various content selection criteria would be applied in different standard or nonstandard presentation and reporting scenarios

**SKILLS DIMENSION**

In addition to the Learner-Beginner skills, **be able to do the following:**

- Demonstrate the practical application of content selection in different standard and nonstandard report situations

### Independent Practitioner

**KNOWLEDGE DIMENSION**

In addition to the Novice Practitioner knowledge, **be able to describe and explain:**

- The criteria and/or principles for assessing the quality of a presentation or report
- How to examine data integrity in preparing and presenting data (that is recognize bias, data fabrication, manipulation, misrepresentation or falsification)
- How to examine data presentation and reporting (that is identify and explain weaknesses in the content, structure and format that could result in suboptimal understanding of the conclusions being presented)

**SKILLS DIMENSION**

In addition to the Novice Practitioner skills, **be able to do the following:**

- Apply quality assessment criteria and principles in practice to produce clear, logical and relevant structures for presentations and reports
- Select the most relevant and appropriate material to be used (and omitted) to present findings and conclusions explaining the rationale for selected material to different target audiences
- Demonstrate the use and application of data integrity principles (existence of bias, data fabrication, data manipulation, misrepresentation or falsification) in the preparation of reports and presentations
- Deliver expertly and confidently – based on thorough preparation – data integrity and subject knowledge
Knowledge Dimension
In addition to the Independent Practitioner knowledge, be able to describe and explain:

- Data visualization tools, what they are and how they are variously used for different effects
- The key information/conclusion requirements for different types of audiences: higher-level manager for decision-making, public audiences, technical experts, field workers or implementation staff at intermediate and lower levels

Skills Dimension
In addition to the Independent Practitioner skills, be able to do the following:

- Prepare, structure and draft high-quality output in a timely manner
- Assess a presentation/report and identify weaknesses in the content, structure and format of the presentation/report that could result in suboptimal understanding of the conclusions being presented
- Use data visualization techniques to present data, analysis and actionable insights
- Demonstrate the presence of data integrity in the presentation of data
- Formulate and present information/conclusion requirements appropriately for different types of audiences: higher-level manager for decision-making, public audiences, technical experts, field workers or implementation staff at intermediate and lower levels
## Domain 4: Transfer to action recommendations

### Learner-Beginner

**KNOWLEDGE DIMENSION**

Be able to describe and explain:

- The core concepts of how data can be used to form actionable insights such as the purpose and principles underpinning the Knowledge to Action (KTA) Framework
- Explain the meaning of each element of the SMART acronym (that is specific, measurable, achievable, relevant, time-bound)

**SKILLS DIMENSION**

Be able to do the following:

- Apply the core concepts (purpose and underpinning) principles of how data can be used to form actionable insights using the criteria of the KTA Framework

### Novice Practitioner

**KNOWLEDGE DIMENSION**

In addition to the Learner-Beginner knowledge, be able to describe and explain:

- How data are used to form actionable insights

**SKILLS DIMENSION**

In addition to the Learner-Beginner skills, be able to do the following:

- Apply the SMART criteria framework to the data to formulate actionable recommendations

### Independent Practitioner

**KNOWLEDGE DIMENSION**

In addition to the Novice Practitioner knowledge, be able to describe and explain:

- The purpose and principles underpinning the KTA Framework in the context of health service delivery systems, and how KTA applies to your data
- How to apply SMART criteria to research and/or recommendations

**SKILLS DIMENSION**

In addition to the Novice Practitioner skills, be able to do the following:

- Develop various strategies for applying actionable insights using the KTA Framework (for example, drawing on best practice examples from other KTA frameworks)
**KNOWLEDGE DIMENSION**

In addition to the Independent Practitioner knowledge, be able to describe and explain:

- How to transition analysis to actionable insights with recommendations as an approach
- How to examine a set of KTA recommendations to identify possible weaknesses to prevent suboptimal recommendations being presented
- How to develop actionable SMART recommendations as a basis for both short and longer-term action plans for different stakeholders
- How to identify potential risks, limitations and challenges of each of the recommendations
- How data will enable future planning for activities and/or priorities in the health-care sector
- How to draw on best practice examples from other sources and to incorporate lessons from those examples to enhance current recommendations being made

**SKILLS DIMENSION**

In addition to the Independent Practitioner skills, be able to do the following:

- Transition analysis to actionable insights with formulation of recommendations as an approach
- Examine a set of data-insights and identify any weaknesses that could result in suboptimal recommendations being presented
- Formulate and communicate a plan for activities and/or priorities in the health-care sector
- Develop actionable (SMART) recommendations as a basis for both short and longer-term action plans for different stakeholder segments
- Draw on best-practice examples from other sources and incorporate lessons from those examples to enhance current recommendations being made
Attitudes

CURIOSITY

PROFESSIONALISM

COLLABORATION

COMMITMENT

COMPLIANCE
Beyond knowledge and skills, behaviours, and mindsets which create an environment characterized by unity of purpose, integrity, respect and the pursuit of excellence are essential for long-term sustainable capacity-building. The Attitudes dimension of the Framework describes four attitudinal domains which direct the behaviour of employees towards professionalism.

<table>
<thead>
<tr>
<th>Domains</th>
<th>Demonstrated by the following observable behaviours while performing duties</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Professionalism</strong></td>
<td>The attitudes, values, conduct and qualities that characterize or hallmark a professional person. Through their actions, professionals consistently demonstrate the qualities of competence, reliability, trustworthiness and integrity in their field of expertise. Being professional means keeping commitments, delivering high-quality work and consistently doing what it takes to demonstrate that you are a team player – being reliable, respectful and competent.</td>
</tr>
</tbody>
</table>
| **Curiosity**    | ● Continuously seeks to acquire new knowledge and skills  
● Wants to know how things work and why things happen the way they do  
● Listens and observes attentively, questions intelligently and learns from experience and from others  
● Open to feedback and self-reflects on actions  
● Seeks to understand the essentials of a new task or role  
● Demonstrates a strong appetite for learning and development  
● Strives to refine existing or acquire new skills |
| **Commitment**   | ● Reliable and dependable  
● Takes responsibility for own actions and holds themself accountable for results  
● Consistently delivers high-quality work output  
● Perseveres to deliver results in the face of obstacles and challenges  
● Proactively identifies ongoing professional development needs and close knowledge/skills gaps for self and others |
| **Collaboration**| ● Communicates and collaborates openly with colleagues and partners  
● Displays empathic behaviour towards others, respectful of differences in cultural and beliefs of others  
● Displays a willingness to help others in need  
● Listens actively and questions respectfully |
| **Compliance**   | ● Engages with colleagues and wider stakeholders in honesty, sincerity and good faith  
● Maintains neutrality from influences and pressures  
● Acts in the best interests of the organization  
● Works diligently and takes pride in own work  
● Follows company policies and complies with procedures  
● Maintains excellent up-to-date work records, ensuring they are current, complete and accurate  
● Maintains confidentiality and integrity  
● Never takes advantage of position for personal gain  
● Declares conflicts of interest, real or perceived, immediately |
3. Implementation in the field

Using this Data Management Competency Framework comprehensively and consistently will yield a host of benefits for the HIW. The key to achieving these benefits is effective and efficient implementation and rollout of the Framework across the organization and its different levels.

In the Introduction chapter, the different target audience segments were identified. Those that are essential as “allies” for the successful adoption and rollout of the Framework were spotlighted. Local ownership of and responsibility for implementation is critical. Local management should embrace the concept and put in place the necessary oversight structures and monitoring procedures to ensure successful implementation.

To this end, it is recommended that the seven-step implementation plan (Fig. 3), which has been designed as a consequence of the testing and piloting process, be used to guide in-country rollout. However, Member States may adjust the order of the steps based on the country context, such as Papua New Guinea case story on page 55.

A local, cross-departmental/divisional project management team, which should include management from the department of human resources and development, should be created to oversee and coordinate the implementation plan.
Fig. 3. The seven-step implementation plan illustrated

1. Mapping (health information) institutional, organizational and individual role architecture
2. Identify and map jobs/roles and determine their data competency profiles, based on current responsibilities
3. Identify capacity gaps through individual employee competency assessments
4. Bring all together through cross-divisional (multidisciplinary) collaboration
5. Determine competency development priorities based on gaps identified in Step 3 and discussions at Step 4
6. Identify and assess all available and relevant resources for better mobilization
7. Develop capacity-building plans for immediate, medium and longer term needs
Step 1: This step involves mapping the in-country health institutions and workforce architecture. It involves three levels of mapping or analysis.

- **Tier 1: Institutional mapping** is focused on getting a "meta" level picture and understanding of the institutional architecture of the in-country health system – see example in Annex 3. This level involves identifying the various and different organizational entities operating within the health sector, and mapping their structures at both national and subnational levels.

- **Tier 2: Organizational mapping** is focused on a more detailed understanding of the internal composition and structure (organogram) of the organizations involved, documenting the different departments, their functional responsibilities, employee numbers, and the degree of engagement that the department has with health data.

- **Tier 3: Individual job role mapping** is focused on the employee level – identifying specific data management roles, their responsibilities and their levels of engagement with health data.

Step 2: This step involves drafting the data competency profiles. This is done by documenting the specific responsibilities of the roles and from that determining the competency areas and competency domains that are necessary to fulfil the responsibilities. Once these have been determined, the next step is to decide the level of proficiency necessary (from learner-beginner to master practitioner) to fulfil the responsibilities to a level of acceptable professional standard (see Annex 1).

Step 3: This is the gap-analysis step and involves a dialogue between employee and line manager whereby both parties discuss (and agree) the actual competency profile of the employee, which is then matched against the (predetermined) required competency profile (Step 2), from which the competency gaps will become evident and visible.

Step 4: This step involves cross-divisional collaboration whereby a cross-functional, multidisciplinary oversight team convenes to assess and evaluate the aggregated gap results and discusses these findings in the context of emerging competency gaps, strategic objectives and resources availability. This coordination team should typically consist of technical line management, the HR department and the training/development department (if separate from HR) and should be chaired by a top management representative under the coordination of a separately appointed team coordinator.

Step 5: This step is about making decisions on training/development priorities – immediate, medium and longer term – taking account of the identified competency gaps and overall strategic objectives. The key criteria for prioritizing response actions should be which competency gaps currently present the greatest risks and threats to the achievement of objectives (department, division, work unit) if left unaddressed. This does not necessarily mean the competency with the biggest gap numbers, but rather the competency gaps that present the biggest threats to achieving institutional objectives.

Step 6: This step involves a discussion on available (usually limited) resources and the most effective allocation of these resources. Resource considerations include the availability of people to lead, manage and deliver the development plan, the funding available, other (ongoing) activities within the organization that require resourcing and other planned activities and training.

Step 7: Once the priorities have been agreed and resource allocation decisions made, the final step is to draft practical and realistic development action plans to address immediate (urgent) gaps and medium-term needs. Training and development plans should be:

- Comprehensive (addressing all significant issues identified);
- Appropriate (delivering the right content to the right people);
- Consistent (applied equally, fairly and transparently); and
- Coordinated (with the right people doing the right things at the right time).

Longer-term development plans do not require detailed actions at this stage as events and priorities change over time. Nevertheless, maintaining the longer-term, sustainable capacity-building perspective is essential and should not be lost sight of. It is this "horizon focus" that provides the "north star" compass-bearing for capacity-building initiatives, which should be focused on building breadth, depth and resilience into the health information workforce over time.
Case story: Papua New Guinea implementation

In order to implement its National Health Plan 2021–2030 and associated monitoring and evaluation (M&E) plan, Papua New Guinea designed a pathway to establish an appropriately skilled HIW that could meet both current and future needs. In collaboration with WHO, Papua New Guinea commenced at Step 2 of the WHO-recommended implementation process to define the competencies of different types of health information roles: national M&E officers, provincial health information officers and medical records officers (Fig. 4). The following steps were taken:

1. Relevant job descriptions were collected and reviewed.
2. A series of interviews were conducted with different national M&E officers, provincial health information officers and medical records officers to evaluate their current roles and responsibilities as well as challenges and expectations.
3. The responsibilities of the three role types and their corresponding competencies were identified. Some critical data management domains, such as data management planning, analysis and interpretation, were not required at less advanced levels. From this analysis the draft data competency profiles were developed.
4. A two-day consultation workshop was organized by involving colleagues from the National Department of Health, including Human Resources, as well as senior management from selected Provincial Health Authorities, health information staff and development partners. The findings and draft data competency profile were reviewed, discussed and further updated.
5. These profiles were presented for final agreement at a workshop two months later involving health information officers from all provinces and senior management from Provincial Health Authorities and the National Department of Health.

The involvement of senior management representatives in the process was critical to ensure understanding of how this work connects to broader organizational objectives, as well as for securing their commitment.

Papua New Guinea will expand the process to cover the HIW at district and health-facility levels and commence Steps 4 and 5 of the WHO procedure to develop capacity-building plans based on competency gaps measured in Step 3.

**Fig. 4.** The Papua New Guinea approach to developing data competency profiles for three HIW roles
References


Annexes

Annex 1

**Indicative example of a data management competency profile for one group of health information officers working at the provincial level**

<table>
<thead>
<tr>
<th>Competence Domain</th>
<th>Learner - Beginner</th>
<th>Novice Practitioner</th>
<th>Independent Practitioner</th>
<th>Master Practitioner</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Generation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data management planning</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data creation</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data collection</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Data maintenance</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><strong>Data Processing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data entry</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Data cleaning</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Data validation</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Data verification</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Data transformation</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><strong>Data Analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Descriptive analysis</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Diagnostic analysis</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Predictive analysis</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Prescriptive analysis</td>
<td></td>
<td></td>
<td>Not a required competency for this role</td>
<td></td>
</tr>
<tr>
<td><strong>Data Usage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data interpretation</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Data Integration</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data reporting</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Transfer to action recommendations</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>
Annex 2

Competency (Proficiency) levels

In any professional activity there are varying degrees of competency. In this Data Management Competency Framework, we have categorized competency into a hierarchy of four levels ranging from learner-beginner to master practitioner as described in the table below.

**Taxonomy of proficiency levels**

<table>
<thead>
<tr>
<th>Level</th>
<th>Classification</th>
<th>Proficiency level descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Learner-Beginner</td>
<td>A graduate-level employee who has just entered the health information workforce and appointed to their first role. Has a level of theoretical knowledge and some practical experience, but this has been acquired in a “controlled-learning” college setting. Because of this, practically oriented work and task-completion experience is very limited, and the employee requires supervised, delegated, on-the-job incremental tasks with regular coaching and feedback.</td>
</tr>
<tr>
<td>2</td>
<td>Novice Practitioner</td>
<td>Has been in the role for one year or more and is developing competency across relevant role domains but continues to work under supervision and oversight on all but most basic delegated professional tasks.</td>
</tr>
<tr>
<td>3</td>
<td>Independent Practitioner</td>
<td>Is independently and reliably competent in the specific domain. Rarely is there a need to refer for advice or guidance – only in unique or exceptional situations</td>
</tr>
<tr>
<td>4</td>
<td>Master Practitioner</td>
<td>Has achieved mastery in the domain. Is an acknowledged expert and expands the boundaries of domain knowledge through research, experimentation and innovation. Is recognized as a trainer and mentor to juniors and is sought after by peers and others for technical advice and guidance.</td>
</tr>
</tbody>
</table>
Three levels of Health-Data engagement:

- Department/Division/Individual is an intensive user of health-related data
- Department/Division/Individual is a moderate user of health-related data
- Department/Division/Individual is a low user of health-related data

Legend and Notes:

1. The alphanumeric characters are used to represent “real world” Directorates/Divisions/Departments and Units. The count, placement and health-data engagement classifications should be interpreted as illustrative rather than prescriptive, indicating the potential for varying structural elements and health-data engagement levels (for example, the Procurement Department might have low engagement while the Surveillance Department might have intensive engagement with health data).

2. The number, distribution and health-data engagement classification of employees under each unit in Tier 3 are similarly illustrative in use to demonstrate the potential for different levels of health-data engagement within a common unit (for example, a statistician working in a particular unit might have intensive engagement, while the team manager of the unit might have moderate engagement).