



**Provincial Department of Health  
Data Governance Framework**



## Contents

Contents .....	1
Acronyms .....	3
Foreword by the Head of Department .....	3
Acknowledgments .....	3
Goals and Objectives .....	4
1. Relevant legislation .....	4
2. Key principles .....	5
Framework Focal Areas .....	6
1. Identify key data governance issues (phased approach) .....	6
Phase I .....	7
Phase II .....	7
Phase III .....	7
2. Obtain executive management sponsorship .....	7
3. Conduct a maturity assessment .....	8
4. Build a provincial road map .....	8
5. Establish a provincial organizational blueprint .....	9
6. Maintain and enhance the provincial and national data dictionaries .....	9
7. Understand the data .....	9
8. Create and maintain a metadata repository .....	10
9. Define the metrics .....	10
10. Perform Master Data Management governance .....	11
Governance structures .....	11
Standards and compliance .....	11
Managing data quality .....	12
Implementation of Master Data Management .....	12
11. Perform analytics governance .....	12
12. Manage data security and privacy .....	13
13. Govern the life cycle of information .....	14
14. Measure results .....	14
Resource List .....	15
Appendix A. Proposed Organizational Structures .....	17
1. Funding mechanism .....	17
2. Executive level .....	17



3. Data governance council .....	18
4. Data stewards.....	21
5. Operational representatives .....	21
Appendix B. Maturity Assessment Matrix .....	22
Appendix C. Data Profiling: Illustrative Issue Detection .....	23
Appendix D. WHO Data Principles for Data Governance Frameworks (2020) .....	24
Appendix E. IBM Master Data Management Implementation Guidance (2010) .....	26

---



## Acronyms

Acronym	Definition
<b>CIO</b>	Chief Information Officer (Provincial)
<b>DHMIS</b>	District Health Management Information Systems
<b>DOH</b>	Department of Health
<b>EIM</b>	Enterprise Information Management
<b>HIE</b>	Health Information Exchange
<b>IBM</b>	International Business Machines
<b>ICT</b>	Information and Communications Technology
<b>KPI</b>	Key Performance Indicator
<b>MDM</b>	Master Data Management
<b>NHISSA</b>	National Health Information System South Africa
<b>PHDGF</b>	Provincial Health Data Governance Framework
<b>PoPI</b>	Protection of Personal Information Act (2021)
<b>SASQAF</b>	South African Statistical Quality Assessment Framework
<b>SOP</b>	Standard Operating Procedure
<b>WHO</b>	World Health Organization

## Foreword by the Head of Department

*[To be obtained from the relevant Provincial Department of Health's Media and Communication Branch once the framework is formally signed off]*

## Acknowledgments

This framework was authored by Anzél Schönfeldt of PATH/Digital Square in fulfillment of a grant provided by the Bill and Melinda Gates Foundation. The framework was developed in response to a technical support request by the Gauteng Provincial Department of Health.

*[Additional acknowledgements to be obtained from the Provincial Department of Health's Media and Communication Branch once the framework is customized and approved]*



## Goals and Objectives

*This framework was developed based on various data governance best practices and is not intended to be used as is; it requires adaptation to provincially specific context(s) in South Africa.*

The goal of the Provincial Health Data Governance Framework (PHDGF) is to coordinate, support, and improve service-delivery outcomes through digital transformation, resulting in improved data quality and use at all levels of the data management system. Effective digital and data governance will enhance provincial digital health initiatives and ensure that health information is well managed, used, and protected in line with national digital health strategies and community expectations. The framework streamlines the implementation of health information systems, allows for the establishment of relevant governance-related roles and structures, and ensures appropriate access to health data, information security, and standardization. Building on the District Health Management Information Systems (DHMIS) Policy of 2011, relevant roles and responsibilities are defined to ensure effective and consistent digital health management across district and health facility data management levels.

### 1. Relevant legislation

It is recommended that the following key strategies, policies, and legislative documents be read in conjunction with the provincial data governance framework:

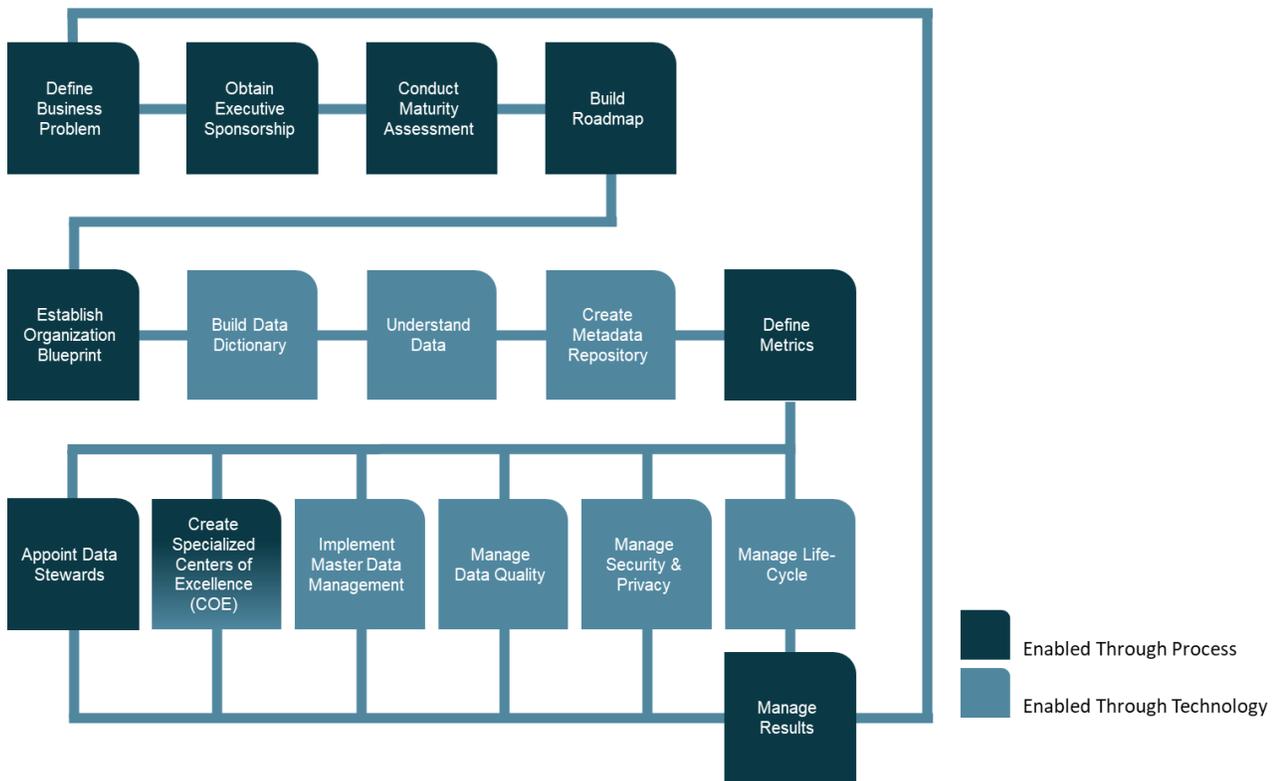
- National Health Act, 2003 (Act No. 61 of 2003).
- State Information Technology Act, 1998 (Act No. 88 of 1998).
- Public Finance Management Act of 1999.
- Statistics Act (Act No. 6 of 1999).
- Promotion of Access to Information Act, 2000 (Act No. 2 of 2000).
- South African National Health Strategy (2015–2020).
- South African eHealth Strategy (2019–2024).
- Health Information System Strategic Plan (internal).
- DHMIS Policy (2011) and associated standard operating procedures (SOPs).
- South African Health Normative Standards Framework, 2<sup>nd</sup> edition (2021).
- Draft of National Data and Cloud Policy (2021).
- Protection of Personal Information (PoPI) Act (2021).
- South African Statistical Quality Assessment Framework (SASQAF), 2<sup>nd</sup> edition.



## 2. Key principles

To ensure successful implementation of the national Health Information System Strategic Plan and associated health policies and strategies over the next several years, a governance structure for digital health and data needs to be formally established. This can be achieved through the alignment and coordination of existing or newly constituted provincial data structures to fulfill specific goals against the PHDGF. The PHDGF model is based on selected elements of the International Business Machines (IBM) Data Governance Unified Process (2010), illustrated in Figure 1 (below), and is augmented by the World Health Organization (WHO) data governance framework principles (2020).

Figure 1: The IBM Data Governance Unified Process.





## Framework Focal Areas

The following IBM framework elements were selected for PHDGF inclusion:

1. Identify key data governance issues (phased approach).
2. Obtain executive management sponsorship.
3. Conduct a maturity assessment.
4. Build a provincial road map.
5. Establish a provincial organizational blueprint.
6. Maintain and enhance the provincial and national data dictionaries.
7. Understand the data.
8. Create and maintain a metadata repository.
9. Define the metrics.
10. Perform Master Data Management governance
11. Perform analytics governance.
12. Manage data security and privacy.
13. Govern the life cycle of information.
14. Measure results.

It is envisioned that the improvement, and in some cases establishment, of these data governance framework components can be developed and deployed concurrently and will likely follow a nonlinear implementation approach based on provincial priorities. Each of these components is detailed below.

### **1. Identify key data governance issues (phased approach)**

For the purposes of this framework, the primary scope is that of data quality improvement to manage risk.

As the project has yet to define its different management structures to implement both the project and its related data governance plan, this document will include suggestions on organizational structures, decision rights, and accountabilities of individuals and groups. As the project is only in its initial start-up phase, it is also recommended that some phases/elements of the organizational process be applied at different stages of system maturity, as detailed in the different sections below.



### ***Phase I***

The first phase comprises the following steps:

- Establishment of data governance management infrastructure and data standards/rules.
- Management of data privacy and security with the provincial Health Information Exchange (HIE)/data repositories, where applicable.
- Monitoring and enforcement to improve data quality and compliance.

### ***Phase II***

The second phase comprises the following steps:

- Conflict resolution.
- Defining of business rules for ongoing governance.
- Strengthening of governance infrastructure and technology and support to data management levels.
- Data dictionary maintenance, standard metadata management, and policy enforcement.

### ***Phase III***

Finally, the third phase comprises the following steps:

- Architecture integration and analysis.
- Management alignment and management support.
- Maturity assessments/risk modeling.
- Dissemination of best practices.

For the purpose of this document, it is assumed that there will be a centralized data governance project office outfitted with minimum administrative support. This office will house the metadata repository/provincial data dictionary instance and centralized provincial HIE databases.

## **2. Obtain executive management sponsorship**

Overall ownership of the provincial data governance drive will belong to the Chief Information Officer (CIO). Executive and senior provincial management must support, sponsor, and understand the activities of the information governance team.

A data governance council will be established by means of a formal terms of reference. This council will lead and take ownership of the provincial data governance processes. Overall ownership of the provincial data governance drive will belong to the CIO, who will function as Chair for the data governance council. The data governance council will be representative of the key functional project areas, including information and communications technology (ICT), health, and executive management.

The CIO is responsible for reporting provincial progress to the National Health Information System South Africa (NHISSA) Committee, which holds the national mandate for health data governance. NHISSA will, in turn, as mandated by its terms of reference, report to the National Health Council Technical Advisory Committee and National Health Council, which are chaired by the National Health Minister.



Alignment with nationally owned data management processes and structures is key in engaging executive management sponsorship, as outlined in WHO's third data governance framework principle (health information systems capacity support). Regular updates between the provincial and national health authorities, as well as between districts within the province, are crucial in championing and obtaining buy-in for any data governance-related activities.

Appendix A outlines a proposed organizational structure, including funding mechanism, executive level, data governance council, data stewards, and other operational representatives.

### **3. Conduct a maturity assessment**

The effectiveness of a data governance program will be assessed against a maturity model. It is advised that the province conduct a maturity assessment at start-up. Appendix B provides a sample of some preexisting maturity assessment models that can either be adapted to fit local context or be used as an example in the development of a new one.

Regular maturity assessments and risk modeling will be instituted during Phase III of the provincial data governance framework, while Phase I and Phase II focus on getting the data governance basics in place. These assessments will be conducted on a semiannual basis, depending on funding availability. Once in effect, it is recommended that the data governance council be responsible for defining the organizational scope of the assessment and associated time frames and identifying the data governance categories to be assessed.

Designated data stewards, in conjunction with the data governance council, will be tasked with identifying participants for the data governance maturity assessment. The data governance council will have overall responsibility for the implementation and dissemination of maturity assessment results to all stakeholders involved in the project and to share the results of the assessment with senior executives.

### **4. Build a provincial road map**

Once a data governance maturity assessment has been conducted, the province will use the summarized results to develop a road map for progressing from the current state of health data and data management systems to the desired future state.

Key processes, people, and technology necessary to bridge the gaps highlighted in the maturity assessment need to be listed and a road map created based on the prioritization of key activities. As a maturity assessment will only be implemented in framework implementation Year 3, it is recommended that the top drivers of data governance, as prioritized in the Identify key data governance issues section, be used to develop an interim road map to improve data governance.

Data quality improvement issues and root causes identified during routine data quality audits by the province, as well as formal results from the Auditor General of South Africa's data management system audits, should assist in putting together an initial baseline for areas of improvement. It is strongly recommended that the road map is costed and presented to the South African National Treasury for central funding allocation in line with the Public Finance Management Act of 1999.



Members from the data governance council will outline the road map, either directly or through the co-option of designated subject-matter experts, while data stewards will take the primary responsibility for socializing and rolling out implementation of the road map across the provincial data management system.

All legislation and policy listed under the [Relevant legislation](#) section of this document will apply to provincial road map development.

## 5. Establish a provincial organizational blueprint

The province needs to establish an organizational blueprint by building a data governance charter that governs its operations. It is advised that a data governance council (as outlined in [Appendix A](#)) be established first and then tasked with outlining and documenting the data governance charter. Once the data governance council has been established, the next level of data governance champions, the data stewards, can be appointed. Ideally, different levels of data stewards are to be appointed to fulfill different roles, as outlined in [Appendix A](#), where practically feasible. Data stewards will informally have accountability for the data they are defining, producing, and using to complete their job or function and should work closely with district, sub-district, and facility information officers during the execution of their duties.

Data governance policies, procedures, and actions need to be monitored constantly and adjusted where necessary, necessitating regular communication and feedback between stakeholders. It is recommended that the data governance council meet on a monthly basis and that the data stewards and operational representatives meet on a fortnightly basis. As the data management system and related data governance processes mature, meetings can respectively start taking place on a quarterly and monthly basis.

## 6. Maintain and enhance the provincial and national data dictionaries

The provincial data dictionary, linked to the provincial HIE, will be maintained in line with prescripts set out for the National Department of Health (DOH) data dictionary, including change control. Any changes to the provincial data dictionary that impact the national dictionary will be communicated to NHISSA by the provincial data governance council. Data stewards should be tasked with ensuring the correctness of data definitions in the data dictionary/metadata repository within their programmatic areas.

## 7. Understand the data

In support of data and health information systems capacity, as outlined by WHO, the Provincial DOH needs to source relevant data analytics training to empower health programmer-specific data stewards to engage with their data, first in conjunction with the provincial monitoring and evaluation team, and later independently.

General supervisory data feedback sessions between Provincial, District, and Sub-district Information Officers, as mandated in the DHMIS policy and SOPs, should include a standing agenda item on data interpretation and improvement. The province should investigate eLearning platforms as a method of content delivery to develop necessary skill levels. National Indicator Data Set training should be conducted at district and sub-district levels on an annual basis to ensure indicators are correctly operationalized and recorded.

Data stewards directly attached to the department's information technology and governance units are to participate in data-profiling exercises annually in order to understand each health data source within the HIE's scope, given the disparate nature of health information systems specifically and relational databases in general. An incomplete



list of data problems that could be explored through data profiling is contained in [Appendix C. Data Profiling: Illustrative Issue Detection](#).

Data-profiling activities should be conducted by the health program–specific data stewards, which will enable the province to develop a complete assessment of the scope and nature of data quality issues.

## 8. Create and maintain a metadata repository

The data governance council will take ownership of the development of a central metadata repository during Phase II of the data governance plan, soliciting the input of the data stewards and other stakeholders to do so. Metadata from the data dictionary, operational metadata, data rules and standards, and data from the data-profiling exercise can be merged as long as the appropriate data lineage is followed.

Impact analyses should be run off the metadata. The data governance council will be responsible for handling any data conflicts, which will be recorded as such (keeping an audit trail intact) and resolved jointly by the council (responsible for standards) and the data stewards (responsible for change control). The resolution should be documented in the change control system.

## 9. Define the metrics

The data governance council is responsible for establishing a scorecard of key metrics to monitor the performance of the data governance program. The Chief Data Steward is responsible for collating input from the data stewards and reporting on these metrics on a regular basis to the data governance council (monthly during Phase I of the data governance plan and quarterly during Phases II and III). Metrics should be categorized in terms of short-, medium-, and long-term interventions.

Apart from using data obtained from the Data Governance Maturity Model to measure overall data governance success, Subramanian's (2017) functional areas will be used to identify data governance success:

- Improvement in data quality scores.
- Adherence to data management standards and processes.
- Reduction of risk.
- Reduction of data rectification costs.

The data governance council will be responsible for coming up with a set of key performance indicators (KPIs). It is key that all data governance stakeholders understand the overall KPIs for the project.

Key metrics as defined for data quality (e.g., completeness of patient records, timeliness of facility-level data collection and capturing, accuracy of lab results by checking against source patient files/cross-checking against pharmacy records) and as stipulated in the SASQAF, National Indicator Data Set, Provincial Indicator Data Set, and DHMIS policy will be adhered to. As the SASQAF is developed by Statistics South Africa, and as the National Indicator Data Set and DHMIS policy originate from the National DOH, the data governance council will work with the provincial monitoring and evaluation unit to develop any data quality measures related to the Provincial Indicator Data Set where and if needed.



The Chief Data Steward will be responsible for overseeing the socialization of any new or updated metrics across the data management system and report back to the data governance council on an as-needed basis.

## 10. Perform Master Data Management governance

### ***Governance structures***

The executive sponsors for each key health programmatic area (domain) will be the divisional representatives who serve on the data governance council. The first step toward master data governance will be the recruitment and appointment of data stewards per domain (preferably already employed by the province and likely to be at District Information Manager or Provincial Programmatic Assistant Director level). The primary objective of the data stewardship role will be to synchronize data collection processes, reduce data redundancy, and increase data accessibility in a systematic manner (i.e., to exercise quality control of data). The data stewards also will be responsible for defining, gathering, and reporting on key metrics relating to their data domains on a monthly basis. A full list of their duties is available in [Appendix A](#).

The executive sponsors will appoint data stewards who will be responsible for data quality on a day-to-day basis. The data stewards will elect a representative to function as the Chief Data Steward. Each domain will have one data steward, while external stakeholders/data providers will each be able to send one data steward representative to be part of the HIE data governance process. As the project expands, each new representative/feeder organization will be able to send a representative to participate in the data steward meetings. It is recognized that some roles and responsibilities may overlap, given challenges with the DOH's staffing establishment and high vacancy rates. Where possible, these overlaps need to be documented and used as motivation for the National Treasury and the provincial Chief Financial Officer to create relevant positions not already contained in the organizational structure.

The data governance council will oversee the data stewardship program and liaise with all data stewards via the Chief Data Steward, who is automatically a member of the data governance council. The provincial executive will review and underwrite/sign off on all recommendations and reports that are issued by the data governance council. The data governance council will have the power to refer any deadlocked complaints processes/noncompliance to the executive for resolution, as outlined in [Appendix A](#).

### ***Standards and compliance***

The data governance council will be responsible for the adoption of all standards in the HIE and ensure that all provinces are compliant with the Health Normative Standards Framework (2021) requirements, as set out by the Auditor General of South Africa, and any [relevant legislation](#) that pertains to standards and compliance in health informatics space.

The data governance council is responsible for enforcement of and compliance with system rules and standards in the provincial HIE. A designated representative will attend National DOH technical working groups on health standards and compliance.



### ***Managing data quality***

As one of the key focus areas for Phase I of the project's data governance plan, data quality is a very important data governance component. The data stewards are charged with the operationalization of data quality, including but not limited to validation, verification, data profiling, and user support.

The data governance council is responsible for setting the scope of internal data quality audits and developing data quality policy in conjunction with the data stewards and other stakeholders. It is recommended that a baseline data quality measurement be taken during Phase I of the project by means of triangulation of data profiling, a data quality audit (both executed by domain data stewards), and measurement of compliance against the data quality policy by each key division in the project. It is recommended that two such triangulation processes are run during Phase I of the data governance plan; as the data management system matures and additional data governance measures come into play, the frequency of this exercise can decrease to once a year during Phase II and biannually during Phase III, as it is important to monitor data quality over time.

Cross-checks between electronic records (e.g., pharmacy dispensary records versus patient files that indicate prescription of medication) and spot checks/random verification of electronic records against source documents can assist in determining the root cause of data quality issues in the HIE and are already part of routine data quality improvement efforts in the province.

Routine data quality monitoring and data cleaning should be conducted by domain-specific/programmatic area-specific data stewards on a weekly basis. It is also advisable that data quality training be initiated by the data governance council, which holds primary responsibility for data governance capacity building throughout the project's data management levels.

### ***Implementation of Master Data Management***

The data governance council will have to develop a business case for any activities related to the implementation of Master Data Management (MDM) on an as-needed basis and submit a proposed budget to the provincial executive/Chief Financial Officer for approval.

Any future implementation of MDM has to be overseen and negotiated by the data governance council, with assistance from the data stewards when and if necessary. MDM implementation guidance, developed by IBM (2010), is included in Appendix E.

## **11. Perform analytics governance**

The province should continue to generate standard reports for their routine data management (e.g., data quality reports, standard monthly and quarterly reports) by means of a District Health Information System data mart that is user-friendly and accessible to all stakeholders (both internal and external). Provision should be made for dynamic reporting in a dashboard or data mart form, and the data governance council, in conjunction with input from the data stewards and operational representatives, should agree on set parameters for standard reporting and analysis.

## **12. Manage data security and privacy**

For the purpose of this framework, security and privacy of data relate to the protection of data stored via computer, server, or any other form of electronic media. As one of the key data governance priorities for Phase I of the project,



the management of security and privacy in the HIE is of utmost importance due to the sensitive nature of some of the data contained within the system.

The project will have to align with key stakeholders on security and privacy measures. As mentioned under the Standards and compliance subsection, all relevant data and privacy legislation and policies need to be adhered to, with specific reference to the following:

- Promotion of Access to Information Act, 2000 (Act No. 2 of 2000).
- PoPI Act, 2021 (see box below).
- State Information Technology Act, 1998 (Act No. 88 of 1998).

Although not legislated, the principles outlined in Appendix D, which are linked to security and privacy, should be regarded as best practice.

#### Key Principles in the PoPI Act

- |  |  |
|--|--|
| a) Patient consent is given for the specific use of personal information.      | d) Permission is given in writing by each responsible party to operators for specific tasks and access to specific data for those tasks. |
| b) Information is in secure and access-controlled data storage.                | e) Confidentiality of personal information must always be maintained.  |
| c) User access is limited to only the information required to complete a task. |  |

These criteria ensure the interoperability with other systems and include security features that protect patient health information. During the deployment phase (Phase I), the health provider database (HIE), as outlined in the National Health Insurance architecture and housed at the Council for Scientific and Industrial Research, should be available for use by all entities participating in the implementation plan to identify appropriate security protocols and test the integrity of the data provided.

Although primarily the task of the provincial data governance council to approve relevant security protocols, data stewards will have to tighten database change controls, automate the compliance workflow processes, define and discover sensitive data, classify and tag the data, and encrypt information accordingly.

The province's ICT team will have to provide the privacy, security, authentication, and access controls necessary for data exchange, including potential group privileges, maintenance of trust certificates, and transport security processes and protocols. Data stewards will assist the provincial ICT team in monitoring the HIE for fraud and cyberattacks and put preventative measures in place to ensure the security of the system and confidentiality of sensitive patient information. The data governance council is tasked with developing access, exception, and extrusion policies, as well as preconfigured policy signatures in conjunction with the data stewards.



### **13. Govern the life cycle of information**

The life cycle of information will be addressed during Phase III of the provincial data governance plan; however, given its involved nature, planning around it would need to start in Phase I.

According to the IBM Model (2010), the information life cycle refers to a policy-based, systematic approach to information architecture, classification, collection, archival use, and retention and deletion of data/records. It involves the establishment of an information architecture, baseline database size and storage architecture, business objects, data classification and defined service levels, and content analysis.

The DHMIS SOPs (2011) will form the basis for updated provincial SOP versions around digitization of existing paper-source documents, overall records management policies, data retrieval, and storage and archives (both cloud and non-cloud). The South African National Data and Cloud Policy draft of 2021 will serve as guidance until finalized and published through relevant official government channels.

For the purposes of this data governance plan, it is advised that the project task the data governance council with establishing policies for managing test data and defining policies for legal discovery of electronic documents during Phase I or II of the data governance plan.

### **14. Measure results**

Results need to be measured against the KPIs developed by the data governance council and executed by the data stewards to monitor performance of the data governance program. Progress against these metrics will in all likelihood ensure continued buy-in from both the donors and the executive level. Measurement results also include baseline assessments, data quality audits and self-assessments, and scorecards. It is recommended that results from these measurements be made available in the public domain by means of a website posting to ensure transparency and demonstrate accountability.



## Resource List

CDI Institute. *Corporate Data Governance Best Practices*. Burlingame, CA: CDI Institute; 2007.

CDI Institute. *Corporate Data Governance Best Practices: 2006-07 Scorecards for Data Governance in the Global 5000*. Burlingame, CA: CDI Institute; 2006. [http://www.tcdii.com/PDF/Data\\_Governance\\_white\\_paper\\_-\\_April\\_2006.pdf](http://www.tcdii.com/PDF/Data_Governance_white_paper_-_April_2006.pdf).

Chapin Consulting Group. *Defining Business Metrics That Matter: Improving Business Results*. Lexington, MA: Chapin Consulting Group; 2006.

Federal Student Aid. *Enterprise Data Management Concept of Operations*. Washington, DC: Federal Student Aid; 2007.

Federal Student Aid. *Enterprise Data Management Data Governance Plan*. Washington, DC: Federal Student Aid; 2007. [https://docshare.tips/data-governance-plan\\_5895568ab6d87f7a088b4e3e.html](https://docshare.tips/data-governance-plan_5895568ab6d87f7a088b4e3e.html).

Federal Student Aid. *Draft Enterprise Data Management Data Policies – Final*. Washington, DC: Federal Student Aid; 2007.

International Business Machines (IBM). *The IBM Data Governance Council Maturity Model: Building a Roadmap for Effective Data Governance*. Somers, NY: IBM; 2007. [http://databaser.net/moniwiki/pds/DataWarehouse/leverage\\_wp\\_data\\_gov\\_council\\_maturity\\_model.pdf](http://databaser.net/moniwiki/pds/DataWarehouse/leverage_wp_data_gov_council_maturity_model.pdf).

Karel R. Data governance remains immature: Increase focus on business process to build momentum [article]. *Information Management*. July 1, 2010. <https://www.smartdatacollective.com/governance-people-data/>

National Department of Health (NDOH). *National Digital Health Strategy for South Africa 2019–2024*. Pretoria: NDOH/Republic of South Africa; 2019. <https://www.health.gov.za/wp-content/uploads/2020/11/national-digital-strategy-for-south-africa-2019-2024-b.pdf>.

Seiner RS. Data steward roles and responsibilities [article in a series]. *The Data Administration Newsletter*. July 1, 2005. <https://tdan.com/data-steward-roles-responsibilities/5236>.

Seiner RS. The data stewardship approach to data governance: Chapter 1 [article in a series]. *The Data Administration Newsletter*. January 1, 2006. <https://tdan.com/the-data-stewardship-approach-to-data-governance-chapter-1/5037>.

Soares S. *The IBM Data Governance Unified Process: Driving Business Value with IBM Software and Best Practices*. Ketchum, ID: MC Press Online; 2010.



Statistics South Africa. South African Statistical Quality Assessment Framework (SASQAF). 2<sup>nd</sup> ed. Pretoria: Statistics South Africa; 2010. [http://www.statssa.gov.za/standardisation/SASQAF\\_Edition\\_2.pdf](http://www.statssa.gov.za/standardisation/SASQAF_Edition_2.pdf).

Subramanian S. Simple metrics for a successful data governance [blog post]. *Towards Data Science*. April 21, 2017. <https://towardsdatascience.com/simple-metrics-for-a-successful-data-governance-ea55b1887d46>.

Williams S, Williams N. *The Profit Impact of Business Intelligence*. San Francisco: Morgan Kauffman Publishers; 2006. <https://www.sciencedirect.com/book/9780123724991/the-profit-impact-of-business-intelligence>.

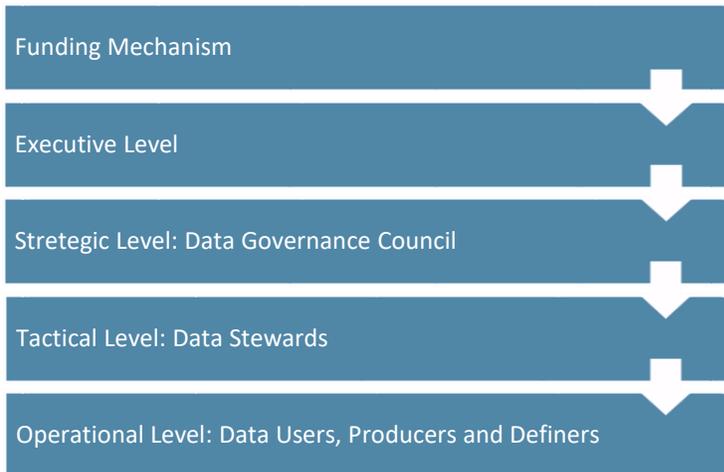
World Health Organization (WHO). Health data as a global public good—a call for health data governance [news article]. World Health Organization. September 29, 2021. <https://www.who.int/news-room/articles-detail/health-data-as-a-global-public-good-a-call-for-health-data-governance-30-september>.



## Appendix A. Proposed Organizational Structures

Figure 1 (below) illustrates a best-practice organizational structure that could be adapted for provincial use.

Figure 1: Overarching structure.



Source: Adapted from Seiner RS. The data stewardship approach to data governance [article in a series]. *The Data Administration Newsletter*. October 2006 to February 2009. <https://tdan.com/the-data-stewardship-approach-to-data-governance-chapter-1/5037>.

### 1. Funding mechanism

Appropriate funding for the data governance initiative could potentially be made available through:

- Fiscal assistance applications via the National Treasury, made available through the office of the Chief Financial Officer within the National Department of Health.
- The contact between any potential donors or National Treasury and the executive in the Chief Financial Officer's office.
- Provincial budget allocations.

### 2. Executive level

**Proposed Chair:** The Provincial Department of Health Chief Information Officer.

**Proposed membership:** The executive committee is the decision-making leadership of provincial data governance activities. Its membership is the current provincial data executive or duly appointed representatives.



**Functions include but are not limited to the following:**

- Nominating a representative or rotating representation for attendance/giving input to data governance council.
- Approving policies and accountabilities as escalated to the executive.
- Signing off on final progress reports.
- Sourcing additional funding/fiscal oversight.
- Providing auditing accountability.

### **3. Data governance council**

**Proposed Chair:** A rotating leadership between representatives of strategic health areas.

**Proposed membership:** One representative from each functional area (Strategic Planning, Data Governance, and Metadata Management; Data Architecture; Data Warehousing; Clinical Data Management; Data Quality and Data Security), Chief Data Steward, representative from the executive level.

It is proposed that the data governance council, once established, maintain and strengthen linkages with the following entities, where applicable:

- Data task force, which consists of invited volunteer representatives from public and private stakeholder organizations, including donor implementing partners, who are contributing time and expertise to better inform the development of data management and governance practices in the province.
- Provincial government staff who conduct work in relation to the provincial data governance framework (e.g., preparing materials, establishing work plans, and coordinating work groups to meet deadlines).
- Governance technical working groups, which consist of volunteers from the data task force who are completing tasks around specific aspects of the data management system.

Joint decision-making/consensus is encouraged; however, decisions pertaining to a specific domain can be referred to the relevant functional area representative (see Table 1, below). In the case of a data governance council deadlock, the Chair heading up the council reserves the right to cast the deciding vote. It is also recommended that the project, via the data governance council, document a Data Constitution and Communications Plan and decision rights per role (e.g., data steward, data owner). Additionally, the process for incorporating new stakeholder data and new representatives/groups into the data governance management framework needs to be outlined and implemented.



**Functions include but are not limited to the following:**

- Enforcing legislation.
- Assigning key roles and responsibilities with each functional area's strengths and areas for improvement in mind.
- Providing conflict resolution by means of a formal complaints procedure with the option of requesting arbitration from the executive for deadlocked matters.
- Choosing representatives from each of the proposed seven key areas given in Table 1 to solicit input from staff involved in each one of the key areas, including but not limited to the Chief Information Officer and internal and external data stewards/stakeholders.
- Providing overall management of the metadata repository/provincial data dictionary.
- Determining the scope of internal provincial data audits and maturity assessments, as well as developing provincial data quality policy.
- Forming subcommittees on compliance, monitoring, and enforcement, and the development, implementation, adaptation, and adoption of new standards and formats in line with Health Normative Standards Framework guidance.
- Disseminating minutes of meetings, new standard operating procedures, policies, procedures, and any other data management–related documentation or information by means of a newsletter to internal and external stakeholders. For this purpose, it is recommended that a Secretariat be established to manage the process.
- Coordinating training and creating awareness around data governance initiatives and the socialization of new governance and data quality improvement measures.
- Relaying, via the Chief Data Steward and executive representative, any data governance decisions to lower levels of the data management systems.
- Liaising with external standards organizations.
- Outlining the data governance road map and overseeing implementation of the Master Data Management guidance.
- As standards custodians, being jointly responsible, with the data stewards (responsible for change control), for data conflict resolutions.
- Developing key performance indicators and establishing a scorecard of key metrics to monitor performance of the data governance program.



Table 1: Proposed functional area structure (non-comprehensive).

<b>Strategic Planning</b>	<b>Data Governance and Metadata Management</b>	<b>Data Architecture</b>	<b>Data Warehousing</b>	<b>Data Quality</b>	<b>Data Security</b>	<b>Clinical Data Management</b>
Perform strategic planning	Define provincial data governance processes	Develop and maintain enterprise data models	Develop and maintain enterprise dimensional data models	Develop and promote data quality best practices	Implement data security and privacy standards in line with national and international best practices and legislation	Promote compliance with health-specific policies, procedures, and standards
Define data policies	Implement data governance processes	Develop and maintain modeling and design standards	Develop data warehouse architecture where applicable	Define data-profiling process	Develop and promote data security and privacy best practices	Promote compliance with health data policies, procedures, and standards in line with relevant national legislation
Promote compliance with data policies, procedures, and standards	Create, capture, and maintain enterprise metadata (data standardization) at a provincial level	Establish and maintain data architecture	Develop data warehouse technical standards and functional specifications where applicable	Manage data quality	Request data security audit from the Auditor General of South Africa	Assist with measuring key health data elements
Disseminate and promote data management	Develop and implement enterprise metadata architecture	Provide database administration support		Monitor and measure data quality		Develop provincial-level health indicators to populate auto-reports (Provincial Indicator Data Set via a District Health Information System dashboard)
Identify and justify resources and budget needs	Create and maintain master data management standards	Assess application integration interface		Conduct data internal quality audits/ assessments		
Monitor program performance		Evaluate tools		Develop data quality dashboards/ reports		

Source: Adapted from the Enterprise Data Management Data Governance Plan (2007).



## 4. Data stewards

There are three types of data stewards:

- Chief Data Steward.
- Data Steward Coordinator.
- Data stewards for programmatic health areas or domains.

Functions include but are not limited to the following:

- Chief Data Steward: reporting to the data governance council on key metrics related to the project data governance and data quality.
- Data Steward Coordinator: acting as contact point for internal data stewards and (data) feeder organizations and compiling relevant sections of the quarterly progress report(s).
- Domain-specific/health program-specific data stewards (i.e., District Information Manager up to provincial specialists): managing change control and arbitrating transformation rules.
- All data stewards: ensuring the correctness of data definitions within their subject areas in the data dictionary/metadata repository.
- As joint or overlapping functions:
  - Overseeing execution of the data governance road map in conjunction with the data governance council, including the championing of the road map to all levels of the data management system.
  - Supporting district and sub-district levels around data quality and operationalization of data quality.
  - Coordinating data verification and profiling, including risk management.
  - As change control custodians, being jointly responsible with the data governance council for resolving data conflicts.

## 5. Operational representatives

Proposed inclusion encompasses but is not limited to the following:

- Users/producers/definers of health data at district and sub-district levels.
- One representative per health facility type per geographical area (e.g., local health clinics, district hospitals, mobile clinics, laboratories, etc.).
- Public health officials/authorities.

**Function:** Process data owners who define/request amendments either to improve processes or to adapt them as the data management system matures.



## Appendix B. Maturity Assessment Matrix

The Maturity Assessment Matrix (below) illustrates a comparison of different maturity models currently in use in the field, which can be adapted for use in a provincial context.

DataFlux	Gartner EIM Maturity Model	EWSolutions
<p>Has logical levels of maturity and dimensions with detailed descriptions.</p> <p>Acknowledges that there is not a single path to reaching a higher level of data governance.</p> <p>Is comprehensive and likely to be adopted with minimum customization in any current context.</p> <p>Has a proven track record of understanding information quality. Seems practical and easy to use; addresses technology adoption and business capabilities that are key to mature enterprise architecture.</p>	<p>Uses “action item” concept.</p> <p>Is a program that evolves over time rather than seen as a single project?</p> <p>Has a concept of managing information as an asset.</p> <p>Is tied to Gartner’s specific definition of EIM.</p> <p>Uses terminology that might “black box” the model for non-information technology staff across the enterprise (e.g., “semantic reconciliation,” “unified content”); might be too complex for non-information technology staff to buy into.</p>	<p>Seems to be too process driven; does not have the detail that, for instance, the Microsoft or DataFlux models have.</p>
IBM	Oracle	Knowledge Statistics
<p>Is detailed.</p> <p>Has a Data Governance Maturity Model that is designed with input from a council of 55 organizations;<sup>1</sup> has group consensus and is widely demonstrated across organizations to work.</p> <p>Has continuous quality improvement features (one of the key principles of quality improvement).</p> <p>Is comprehensive and covers a wide range of domains (11) and the <i>interaction</i> between their four major groupings.</p>	<p>Aligns people, processes, and technology.</p> <p>Talks about feedback loops that are usually missing in other literature.</p> <p>Features MDM.</p>	<p>Is not detailed enough for provincial use (lack of descriptions leaves the characteristics open to interpretation).</p>
	MDM Institute	
	<p>Is a bit too basic but likely to have its place in less complex settings; does not provide sense of how the levels interact/transition.</p>	

Abbreviations: EIM, Enterprise Information Management; IBM, International Business Machines; MDM, Master Data Management.

<sup>1</sup> IBM Software Group. *The IBM Data Governance Council Maturity Model: Building a Roadmap for Effective Data Governance*. Somers, NY: International Business Machines (IBM) Corporation; 2007. [http://databaser.net/moniwiki/pds/DataWarehouse/leverage\\_wp\\_data\\_gov\\_council\\_maturity\\_model.pdf](http://databaser.net/moniwiki/pds/DataWarehouse/leverage_wp_data_gov_council_maturity_model.pdf).



## Appendix C. Data Profiling: Illustrative Issue Detection

Some of the issues that can be explored via data profiling include but are not limited to the following:

- “Business” rule validation: ensure your data meet organizational standards for data quality and health processes by validating data against standard statistical measures, as well as customized validation rules.
- Relationship discovery: uncover relationships across tables and databases and across different source applications.
- Outlier detection: detect data that fall outside of predetermined limits and gain insight into source data integrity (e.g., when the number of patients seen at a specific health facility during a specific time period for a specific ailment/category exceeds the total head count of the facility).
- Data validation: verify that data in your tables match their appropriate description (e.g., Nevirapine is described as an antiretroviral and not cough medication).
- Pattern analysis: ensure that your data follow standardized patterns to analyze underlying data and build validation rules.
- Data life cycle quality: discover the data lineage and complex transformation logic between sources.
- Statistical analysis: establish trends and commonalities in corporate information and examine numerical trends via mean, median, mode, and standard deviation.
- Data security: identify and protect sensitive patient data inside each data source (e.g., patient diagnosis or lab results).
- Data source management: understand how data overlap across data sources for critical data elements (e.g., the link between tuberculosis and HIV/AIDS data elements).



## Appendix D. WHO Data Principles for Data Governance Frameworks (2020)

WHO Data Governance Framework Principle (adapted for local context)	Relevance to the Provincial Department of Health Data Governance Framework	Potential Data Quality Dimension
<b>1. Data shall be treated as a public good</b>		
<ul style="list-style-type: none"> <li>• Effort should be made to release data publicly and to share when safe and ethical to do so.</li> <li>• Clear guidance is to be provided in situations where legitimate reasons prevent the sharing of data (i.e., possible ways in which the data may be accessed, such as for research purposes). The guidance will be consistent with applicable rules and policies, including data-sharing policies, information disclosure policy, and personal data protection policy.</li> <li>• Informed consent is almost always needed for research. For routine data collection for purposes of public health surveillance, informed consent is not the default and is not always required.</li> <li>• For personal data, the consent of the data subject should be the preferred basis for processing the data.</li> <li>• There should be transparency around how data are collected, used, and shared. Adherence to this principle, therefore, requires that complete metadata about datasets should be released along with the datasets themselves and together with any explanatory notes that may be required to provide context on data provenance, scope and limitations, application and (re-)use, traceability, and sharing.</li> </ul>	<p>Compliance with the Protection of Personal Information Act (2021)</p> <p>Future Compliance with the Draft Data and Cloud Policy (2021)</p> <p>Access to Provincial and National Data Sharing Agreement guidance and application forms</p> <p>Compliance with the promotion of Access to Information Act, 2000 (Act No. 2 of 2000)</p> <p>Compliance with the South African Quality Assessment Framework (2<sup>nd</sup> edition)</p>	<p><b>Accessibility</b> (ease of data attainment from government):</p> <p><i>Numerator</i> – percentage of data user agreements signed in line with the District Health Management Information Systems Policy and associated standard operating procedures;</p> <p><i>Denominator</i> – number of compliant data user agreements received during a set time (e.g., annually)</p> <p><b>Transparency</b></p> <p><b>Data Provenance</b></p> <p><b>Relevancy</b></p>
<b>2. Trust in data shall be upheld</b>		
<ul style="list-style-type: none"> <li>• Data that are shared and placed under the organization’s control shall uphold the trust placed in it by data originators.</li> <li>• No data will be shared outside of context.</li> <li>• A consultation process will be put in place to consult with data originators prior to data sharing.</li> <li>• Data will be securely and confidentially stored and processed.</li> <li>• Data originators will have to confirm that the data have been collected in accordance with applicable national laws, including data protection laws aimed at protecting the confidentiality of identifiable persons.</li> </ul>	<p>Compliance with the Protection of Personal Information Act (2021)</p> <p>Compliance with the South African Quality Assessment Framework (2<sup>nd</sup> edition)</p> <p>Compliance with the promotion of Access to Information Act, 2000 (Act No. 2 of 2000)</p>	<p><b>Integrity:</b> values and related practices that maintain users’ confidence in the provinces’ data and data products</p> <p><b>Data Provenance:</b> data origin (relating to data collection)</p> <p>Could include <b>Data Lineage:</b> data origin, including analytic life cycle</p>



WHO Data Governance Framework Principle (adapted for local context)	Relevance to the Provincial Department of Health Data Governance Framework	Potential Data Quality Dimension
<b>3. Data and health information systems capacity shall be supported</b>		
<p>The province should endeavor to provide data and health information systems capacity at provincial, district, and sub-district levels through:</p> <ul style="list-style-type: none"> <li>• Sharing best practices around data governance, health management information systems, public health statistics, health-related data science, and health data innovation.</li> <li>• Advocating for evidence-based decision-making by focusing on sustainable health information management systems and digital development systems.</li> <li>• Strengthening capacity to collect, analyze, disseminate, and use data to develop and monitor provincial policies and plans.</li> <li>• Aligning with nationally owned monitoring and evaluation processes and structures.</li> <li>• Reducing provincial reporting burdens.</li> </ul>	<p>Provincial, district, and sub-district supervision and data feedback meetings as outlined in the DHMIS Policy and associated standard operating procedures</p> <p>Provincial Indicator Data Set annual and biannual development processes feeding into the National Department of Health data dictionary</p> <p>DHMIS Policy Compliance</p>	<p><b>Accessibility</b></p> <p><b>Transparency</b></p>
<b>4. The province shall be a responsible data manager and steward</b>		
<ul style="list-style-type: none"> <li>• The province will ensure that all data made available to it are processed, maintained, analyzed, disseminated, and used in accordance with national standards and best practices in health data management.</li> <li>• The province shall ensure that all data it produces are of consistently high standards that include transparent audit trails and common reference years, as well as being timely, accurate, and (where technically and legally possible) accessible.</li> </ul>	<p>Compliance with the Protection of Personal Information Act (2021)</p> <p>Future Compliance with the Draft Data and Cloud Policy (2021)</p> <p>Compliance with the South African Quality Assessment Framework (2nd Edition)</p> <p>Compliance with the South African Health Normative Standards Framework (2021)</p>	<p><b>Data Provenance</b></p> <p><b>Timeliness</b></p> <p><b>Accuracy</b></p> <p><b>Accessibility</b></p>
<b>5. The province shall strive to fill public health data gaps</b>		
<p>Where possible and appropriate, the province will use empirical data collection and predictive, transparent, and coherent modeling methods with proven validity to fill any data gaps.</p>		<p><b>Completeness</b></p> <p><b>Accuracy</b></p>

Source: World Health Organization (WHO). World Health Organization Data Principles. Geneva: WHO; 2020. [https://cdn.who.int/media/docs/default-source/world-health-data-platform/who-data-principles-10aug-\(3\)\\_64f62ef2-254b-41f4-9e3e-57ee6330b0bb.pdf?sfvrsn=3d89acf0\\_6](https://cdn.who.int/media/docs/default-source/world-health-data-platform/who-data-principles-10aug-(3)_64f62ef2-254b-41f4-9e3e-57ee6330b0bb.pdf?sfvrsn=3d89acf0_6).

Abbreviations: DHMIS, District Health Management Information Systems; WHO, World Health Organization.



## Appendix E. IBM Master Data Management Implementation Guidance (2010)

IBM recommends the following guidance in terms of Master Data Management implementation:

- Identify the business problem: In the case of the Health Information Exchange (HIE), the data governance plan has already identified and prioritized key areas of concern.
- Define the subject areas of master data: Proposed key master-data entities for the HIE could include the patient, health care facility, physician, pharmacy, lab, and prescriptions.
- Identify the systems and business processes that consume the data: A mapping exercise is recommended to fully understand the feeding systems, as well as processes both internal and external to the project.
- Identify the current data sources: The Master Data Management system and its documentation will be a dynamically changing one as it is envisioned that more data sources and feeder organizations will be added to the HIE in due course. This changing environment should be taken into account when documenting processes and procedures.
- Define the data attributes of the system of record: For example, in the HIE, the attributes for “patient” could include social security number, medical history, address, contact details, and health insurance information.
- Appoint data stewards for each system of record: This section has already been covered in the data governance plan.
- Establish policies for master data governance, which should already be covered in the data governance council’s mandate and should include data matching and validation rules, change control, naming conventions, and identification of specific attributes as sensitive data (e.g., lab results).
- Implement a data steward console for manual intervention and monitoring.
- Manage potential overlay tasks, which occur primarily when data are updated (e.g., a patient with the name of Anzel Schonfeldt with a unique ID of 334 who is then updated to reflect a John Doe with a unique ID of 334). Such instances should be prioritized for immediate investigation and corrective action where necessary.
- Match duplicate suspects from the same source or from multiple sources to create a new master record; it is possible that a patient shows up in two unrelated datasets with exactly the same information, and those datasets should be combined/consolidated into a new record.
- Link related records from multiple sources (e.g., a patient entry for “William Clinton” and another entry for “Bill Clinton” where the other data attributes linked to both these entries are exactly the same and so should be addressed).
- Review duplication of unique identifiers that should be addressed (e.g., if two different patients show up in the HIE with exactly the same social security number).
- Manage data relationships, hierarchies, and groupings.
- Architect the Master Data Management Solution: The project will need to decide whether to follow a transactional, registry, analytical, or hybrid architecture. This decision should be delegated to the data governance council and its representatives.