



SDAIA

الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority

National Data Index

Operational Excellence (OE)

Frequently Asked Questions (FAQs)

Document Control

Version	Revision Date	Contributor	Modification
1.0	March 2024	SDAIA	Final Version
2.0	August 2024	SDAIA	Final Version

Table of Contents

Document Control	2
1. Glossary.....	4
2. Introduction	5
3. Frequently Asked Questions (FAQs)	6

1. Glossary

Term	Description
DMP	Data Marketplace
GSB	Government Service Bus
NDB	National Data Bank
NDC	National Data Catalog
NDI	National Data Index
NDL	National Data Lake
NDMO	National Data Management Office
NIC	National Information Center
ODP	Open Data Platform
OE	Operational Excellence
RDP	Reference Data Platform
SDAIA	Saudi Data and Artificial Intelligence Authority

2. Introduction

The Saudi Data and Artificial Intelligence Authority (SDAIA) is on a mission of developing a data-driven economy and fostering the data literacy level across the government agencies for operational efficiency and decision-making in the Kingdom of Saudi Arabia (KSA). To this end, SDAIA has launched the National Data Index (NDI) framework that aims at measuring the efforts and progress in transforming the data into a vital economic resource for unlocking innovation, driving economic growth and transformation, and improving the national competitiveness in an organized and accelerated manner.

NDI encompasses the National Data Management Office (NDMO) Data Management and Personal Data Protection Framework prescribed domains and is composed of three essential components: Compliance, Maturity, and Operational Excellence (OE). Each government entity will be measured on a regular basis across these three components.

This document is developed specifically to collate all the commonly asked questions pertaining to OE in a single place. This document provides clear and concise answers for these questions to help concerned people obtain further clarity about the OE metrics.

3. Frequently Asked Questions (FAQs)

What is “Operational Excellence (OE)”?

OE is a framework for measuring the government agencies performance in the data management and governance area through the national data platforms developed and managed by SDAIA.

What is the relationship of OE with the National Data Index (NDI)?

NDI consists of three essential components: Compliance, Maturity, and OE.

How many metrics does OE have in total?

OE has 20 metrics spreading across six data management domains as follows:

Data Management Domain	Metric Count
Data Sharing and Interoperability	3
Open Data	2
Data Catalog and Metadata Management	3
Reference and Master Data Management	3
Data Quality	5
Data Operations	4

What do you mean by “National Data Platforms”?

A National Data Platform is a technology platform which government agencies are required to use for data management, governance, and/or sharing of national data assets.

The following platforms are currently considered as National Data Platforms:

- Government Service Bus (GSB): A central platform that enables the government agencies in the Kingdom to integrate and interlink to share standardized transactional data on-demand for facilitating government business processes.
- National Data Lake (NDL): A reliable central repository for preserving, processing, and cleansing national data and then sharing it securely with the beneficiaries to enable them to build decision-support platforms.
- Collaborative Data Labs (CDL): Secure labs that enable employees of government agencies to access the data hosted in the National Data Lake through an environment equipped with the latest data analytics and AI technologies to discover, explore, and analyze the data and generate the required insights and reports.
- Data Marketplace (DMP): A platform that aims at automating the data-sharing processes across the agencies in the Kingdom. It enables them to browse the data-sharing services and subscribe to what suits them in an automated manner in accordance with the national data governance policies.
- National Data Catalog (NDC): A platform that serves as an inventory of the metadata of the government agencies’ systems, with the definitions of their Key Performance Indicators (KPIs) and metrics and the list of business-critical data fields linked with their certified sources, in addition to the data policies and standards.

- Reference Data Platform (RDP): A platform that offers the features to support extensive curation capabilities for standardizing, classifying, and defining the ownership for the reference data at a national-level across the government agencies and ensures completeness, accuracy, and consistency of the available reference data.
- Open Data Portal (ODP): A platform that enables individuals, government, and non-government agencies to publish their open data and make it available to end users, such as entrepreneurs, to enable them to develop innovative products that contribute to building a digital economy in the Kingdom.

How is the OE final score calculated?

The OE metrics' scores have different measurement units (e.g., time and percentage), and hence, to normalize these units, each metric output is mapped to a predefined scale as follows:

Scale Level	Scale Value
Unacceptable	0
Low	1
Fair	2
Good	3
Excellent	4
Leader	5

The OE score of an agency is calculated using the following weighted sum equation:

$$OE\ Score = \sum_{i=1}^{|L|} w_i * s_i$$

where $L = \{m_1, m_2, \dots\}$ is the list of metrics considered for the assessment round, s_i represents the score/scale value of the agency for metric m_i , and w_i represents the weight of that metric.

How many metrics are targeted for the first year/round of the assessment and what is the plan for the remaining metrics?

For the first year, six metrics are targeted which are listed below:

Metric ID	Metric Name	Platform
DSI.OE.02	Systems integrated with NDL	NDL
DO.OE.02	Failed API calls on GSB	GSB
DO.OE.03	Operational issues from entities encountered by NDL	NDL
OD.OE.01	Datasets published in ODP	ODP
MCM.OE.01	Systems cataloged in NDC	NDC
MCM.OE.02	Business attributes defined and linked in NDC	NDC

What are the weights targeted for the first year?

The following table lists the weights for the six metrics targeted for the first year.

Metric ID	Metric Name	Metric Weight
DSI.OE.02	Systems integrated with NDL	0.20
DO.OE.02	Failed API calls on GSB	0.15
DO.OE.03	Operational issues from entities encountered by NDL	0.15
OD.OE.01	Datasets published in ODP	0.20
MCM.OE.01	Systems cataloged in NDC	0.15
MCM.OE.02	Business attributes defined and linked in NDC	0.15

Is there a possibility of adding new domains or metrics in OE in future?

Yes, new domains or metrics will be added in the future based on the arising needs. However, those domains and metrics will be communicated with the agencies in advance to prepare for the evaluation.

Can the weight of a metric be changed?

Yes, based on the priority and importance of the metric, weights for any metric can be changed. This provides flexibility to NDI to cater the high priority domains and metrics. It will also help agencies to focus on the areas of improvement. However, it is important to note

that the weights for the metrics can possibly be changed in the subsequent rounds and will remain unchanged for the current round.

Is there any requirement for submitting proofs or evidences from the agency pertaining to OE metrics?

No, there is no expectation from the agencies to submit any materials. The National Data Platforms shall be used to collect the evidences and required data for progress measurement.

In metric DSI.OE.02, what do you mean by “Systems integrated with NDL”?

This metric will calculate the percentage of the systems integrated with NDL against the total number of systems requested by the NDL team at the beginning of the data integration project. Note that a system will be considered integrated only it adheres to following three conditions:

- Agency and NDL are connected and a staging database were made accessible to NDL through a secure connectivity.
- The integration method shared by the NDL team is followed completely.
- The database tables, which are part of the integration, are refreshed or updated regularly on defined schedule.

Please show an example of how the metric DSI.OE.02 is calculated?

The following example illustrates the method for calculating the metric. Suppose that the NDL team has requested five systems to integrate with and that the integration status of these systems is as follows:

System	Integration Implemented?	Following Integration Method?	Data Updated Regularly?
System 1	Yes	Yes	Yes
System 2	Yes	No	Yes
System 3	Yes	Yes	No
System 4	Yes	Yes	Yes
System 5	No	No	No

Based on the integration status above, only System 1 and System 4 are considered integrated and hence the metric score is calculated as follows:

$$\text{Score} = (\text{Fully Integrated Systems} / \text{Total Number of Systems}) * 100 = 2/5 * 100 = 40\%$$

Scale Interval = Unacceptable

In metric DO.OE.02, what is meant by “Failed API calls”?

Any API call receiving an error code is considered as a failed API call. Any issue pertaining to GSB platform itself or client-side is not part of the API failure calculation.

Please show an example of how the metric DO.OE.02 is calculated?

Suppose that the agency has five APIs published on GSB and that the number of calls and number of failed calls on these APIs are as follows:

API	Number of Calls on the API	Number of Failed Calls the API
API 1	100	3
API 2	80	5
API 3	150	6
API 4	200	7
API 5	30	7

Based on the status above, the metric score is calculated as follows:

$$\text{Score} = (\text{Total Number of Failed Calls on All APIs} / \text{Total Number of Calls on All APIs}) * 100$$
$$= (3+5+6+7+7) / (100+80+150+200+30) * 100 = (28/560) * 100 = 5\%$$

Scale Interval = Unacceptable

In metric DO.OE.03, what do you mean by “Operational Issues”?

Data tables from source systems are fetched on regular intervals. Each time a data acquisition job executes on the defined schedule, it is termed as an execution cycle.

If an execution cycle is triggered and the job is failed, the NDL operations team tries it multiple times to complete the data acquisition. In case two or more errors occur due to an issue from the agency side in that cycle, it is considered as an operational issue. It is important to

note that multiple failures in an execution cycle is counted as only one issue. For example, suppose that the data acquisition job is scheduled for a daily run, and it fails to connect with the agency’s staging database designated for NDL due to database unavailability. The NDL operations team will re-run it again and in case the job fails to complete successfully two or more times, it will be classified as an operational issue and counted as one failure for that execution cycle. Note that a maximum of one operational issue will be counted per each table per day. This metric considers only the issues arising from the agency side only (e.g., database unavailability, altering the data types or column names, missing table).

Please show an example of how the metric DO.OE.03 is calculated?

Suppose that the agency has prepared a staging database containing two tables for NDL, and that the data acquisition jobs are scheduled on a daily basis for these two tables. Below is an example of the execution status for the tables for one month.

Table	Day	Total Execution Tries	Failed Execution Tries	Final Status	Operational Issues Count
Table 1	Day 1	3	3	Failed	1
Table 1	Day 2	2	1	Successful	0
Table 1	Day 3	3	2	Successful	1
Table 1	Day 4 – Day 29	26	0	Successful	0
Table 1	Day 30	3	3	Failed	1
Table 2	Day 1	2	1	Successful	0

Table 2	Day 2	2	1	Successful	0
Table 2	Day 3 - Day 30	28	0	Successful	0

$$\text{Failure Percentage} = \text{Total Failed Pipelines Executions} / \text{Total Pipelines Executions} * 100$$

$$= 3 / 60 * 100 = 5\%$$

Scale Interval = Fair

In metric OD.OE.01, what do you mean by “total number of published datasets”?

The published datasets refer to the datasets that have been uploaded to ODP by the agency and approved by the ODP team for publication.

In metric OD.OE.01, what do you mean by “total number of datasets required to be published in ODP by the agency”?

The total number of required datasets will be communicated to the agency by SDAIA team.

Please show an example of how the metric OD.OE.01 is calculated?

Given that the assessment period of the first NDI round is the first six months of 2024, and assuming that the number of required datasets for this period is 40, and that the agency has published 30 datasets only in the same period, the score is calculated as follows:

$$\text{Score} = \text{Number of Datasets Published in ODP in the Assessment Period} / \text{Total Number of Datasets Required to be Published in the Assessment Period} * 100 = 30 / 40 * 100 = 75\%$$

Scale Interval = Low

In metric MCM.OE.01, what do you mean by “Technical Column”?

A technical column is a type of metadata that describes the technical properties of a database’s physical column, such as the data type, field length, content profiling, and lineage. The following table provides examples of a technical column.

Physical Column Name	Datatype	Storage (Bytes)	Format	Constraint
Dt_of_birth	DATE	1	YYYY-MM-DD	NOT NULL
Blood_grp_cd	TINYINT	4	NONE	> 0

In metric MCM.OE.01, what do you mean by “Critical System”?

Critical system is a type of an application which is essential to operate a business function effectively. A critical system holds the following characteristics:

- It is involved in frequently used high-level reporting and summaries to managers and executives of the agency.
- Its data is considered essential for running business and making decisions by those responsible for the managing the core business functions of the agency.
- It contains high potential value and outcomes for government, industry, or the community.

In metric MCM.OE.01, when is a system considered cataloged?

A system is considered cataloged if it is added in NDC under the system facet along with its sub-systems, and all of its technical metadata is uploaded in NDC.

Please show an example of how the metric MCM.OE.01 is calculated?

Suppose the agency has six critical systems and one of them (i.e., System 5) cannot be cataloged due to technical difficulties in extracting its technical metadata from its own databases. Now, suppose the agency has provided the technical metadata for four of the remaining five systems only (i.e., System 1, System 2, System 4, and System 6).

System	Technical Metadata Provided?
System 1	Yes
System 2	Yes
System 3	No
System 4	Yes
System 5	N/A
System 6	Yes

Based on the status above, the metric score is calculated as follows:

$$\text{Score} = (\text{Number of Systems Cataloged} / \text{Number of Systems Required to be Cataloged}) *$$

$$100 = 4/5 * 100 = 80\%$$

Scale Interval = Fair

In metric MCM.OE.02, what do you mean by “Business Attribute”?

Business attribute is a type of metadata that describes the meaning of a physical column by providing a context with respect to the business function. Business attributes help users find and understand data, as well as govern and manage it effectively. The following table provides examples of a business attribute.

Physical Column Name	Attribute Name	Attribute Definition	Business Constraint
Dt_of_birth	Date of Birth	The date of which the person was born	<ul style="list-style-type: none">• It must be less than the current date• It cannot have a future date
Blood_grp_cd	Blood Group Code	The numerical code assigned to the blood group	<ul style="list-style-type: none">• Codes must be from the defined list of blood groups

In metric MCM.OE.02, what is the way to determine the “Total number of business attributes”?

The number of business attributes required for a system should be at least 10% of the number of technical columns of that system. For next years, this percentage may change due to the scale and complexity of the systems.

Please show an example of how the metric MCM.OE.02 is calculated?

The following example illustrates the method for calculating this metric.

System	Number of Technical Columns	Number of Required Business Attributes	Number of Defined Business Attributes
System 1	270	27	21
System 2	1080	108	92
System 3	360	36	29

The completion percentage of System 1 is calculated as follows = Number of Defined Business Attributes / Number of Required Business Attributes * 100 = 21 / 27 * 100 = 77%

Similarly, the completion percentage of System 2 and System 3 are 85% and 80% respectively. Hence, the overall score of the metric is calculated as the average completion percentages of the three systems as follows:

$$\text{Score} = (77\% + 85\% + 80\%) / 3 = 80.6\%$$

Scale Interval = Good

In metric MCM.OE.02, what is the meaning of “linked in NDC”?

It means a business attribute in NDC must be linked with a unique technical column of the system uploaded in NDC to define a relationship between the business attribute and technical column. The following example shows the linking of a business attribute with a technical column.

Business Attribute	Physical Column
Date of Birth	Dt_of_birth
Blood Group Code	Blood_grp_cd

Note that agencies are not required to link the business attributes with technical columns for the first year.

