



**Ministry of Digital Economy  
And Entrepreneurship**



**OPEN GOVERNMENT DATA**

## **Procedural Framework for Measuring and Evaluating The Quality of Governmental Open Datasets**

**2020**

## Table of Content

PREFACE.....	3
INTRODUCTION: .....	4
THE MECHANISM OF EVALUATING INDICATORS.....	6
REGULATIONS COMPLIANCE .....	8
PRIORITIZATION.....	9
PUBLISHING DATASETS FIRST TIME.....	10
UPDATING PHASE – AT EACH NEW RELEASE .....	16

# Quality Of Governmental Open Datasets

## Preface

The last decade witnessed a remarkable development in programs and initiatives related to open government data, as many countries quickly launched their own initiatives and created portals for the proactive disclosure of their data in an open legal and technical manner, which led to the publication of a greater number of government data globally.

Although the publication of open data by government entities as aims basically at emphasizing the principles of good governance and transparency, in addition to ensuring social participation and providing new opportunities that grow up the national economy through the reuse of data in several ways to achieve social and economic benefits. The rates of usage did not rise to the required levels, which was due to publishing datasets that did not match the real demand for data. In addition to the poor quality of open data on several levels, including technical and legal related aspects and others related to the data content itself.

Government perceives that publishing low quality datasets usually has a negative impact on transparency and building trust with beneficiary, as most of the current methodologies for evaluating the quality of open data lack theoretically and in practice the Universality due to the diversity and difference of datasets. Moreover, most of the evaluation mechanisms focus on evaluating platforms without examining data in all its aspects.

Accordingly, and with the aim of improving the quality of open government data in Jordan, the Ministry of Digital Economy and Entrepreneurship (the Ministry) prepared this procedural framework. This effort has been done in cooperation with the Organization for Economic Cooperation and Development (OECD) and as part of the Joint Committee for Open Government Data mandate, to assist government entities for developing its own procedure to ensure the quality of its datasets through assessing the quality of open government data based on unified criteria.

# Quality Of Governmental Open Datasets

## Introduction:

According to Cabinet Resolution No. (3803), the Jordanian government approved an open government data policy in 2017, which aims to facilitate access to open government data, increase transparency and confidence in government performance and increase the participation of civil society in the policy and decision-making processes, in addition to creating a new resources for entrepreneurs, and encouraging business and creativity and investment in the Kingdom. For achieving the policy objectives, the Ministry issued instructions for publishing "Open Government Data on the open government data platform" issued in 2019 in the Official Gazette No (5561) page (660-663). This instruction aims to govern managing the open data implementation in each government entity through assigning a set of roles and responsibilities to coordinator and lesion officer. Moreover, this instruction clarifies steps to follow in the process of publishing open government data, which formulated to increase the number of published open datasets on the platform.

As one of the main instructions' requirements needed for the publishing of open government data, the government issued the Jordanian open data license on 20/08/2019; the License was circulated it to all government entities to adhere to its content, aiming to promote the use of open government data, and grant beneficiaries the freedom to share, modify, use and re-use the published datasets for any purpose without restrictions, taking into account a number of requirements of privacy, protection and franchise rights as conditions of use.

Under the second pillar of commitment no. (2) Of Jordan's OGP Fourth National Action Plan 2018-2020, which includes the need to develop procedures for classifying, measuring and evaluating the quality of open government data. The Ministry cooperated with the OECD, as part of the joint committee for open government data mandate of preparing this procedural framework, and defining the tasks of data coordinators and liaison officers in government agencies, based on their responsibilities stipulated in the instructions for publishing open government data, to ensure the accuracy and quality of open government data published on the open government data platform, or those that will be published in the future by Governmental entities.

To ensure the legal suitability of open data sets, it is necessary to start studying the legal requirements before the start of the evaluation process according to the stages of the data sets. And to measure the data quality index for each dataset, sub-indicators were measured to qualify open data sets, which are: data relevance, validity, completeness, ease of access, primacy of its source - partiality of data, appropriate timing for publishing it, openness, automated processing and finally feedback. This process goes through measuring twenty-eight sub-indicators distributed into three successive evaluation phases, with no specific timeframe phases. Where the evaluation process begins with the phase of determining the priority data for publication measured through five sub-indicators, then to the phase of "first time publishing" on the platform with a twenty-one indicators, then and there the updated versions of the data set evaluated in this phase. On all the indicators mentioned in the publishing phases for the first time, in addition to three additional indicators, with a twenty-four indicators

## Quality Of Governmental Open Datasets

in this phase. In addition, it worth mentioning here that the evaluation process includes all data, regardless being data set itself or metadata.

This framework designates the responsibilities of evaluating indicators by data coordinators and content link officers in government agencies. Therefore, each government agency must prepare a detailed internal procedure in line with this procedural framework within a year after the issuance of this framework. The institutional performance development department in each entity should prepare internal procedure, who will be assessing the implementation of the internal procedure. The internal procedure will indicate that the sequence through which data coordinator will obtain the data of normal classification, based on the periodic update of the open "Government Data Collection Inventory Form", in addition to the outputs of the Institutional Data Classification and Management Committee found in the Government Data Classification and Management Policy 2020.

In addition, the ministry, in partnership with the stakeholders, will conduct a periodic evaluation of the implementation of this procedural framework in government agencies. Based on that, the ministry will analyze the results and ensure the necessary adjustments, depending on the feedback from the government agencies and the beneficiaries.

# Quality Of Governmental Open Datasets

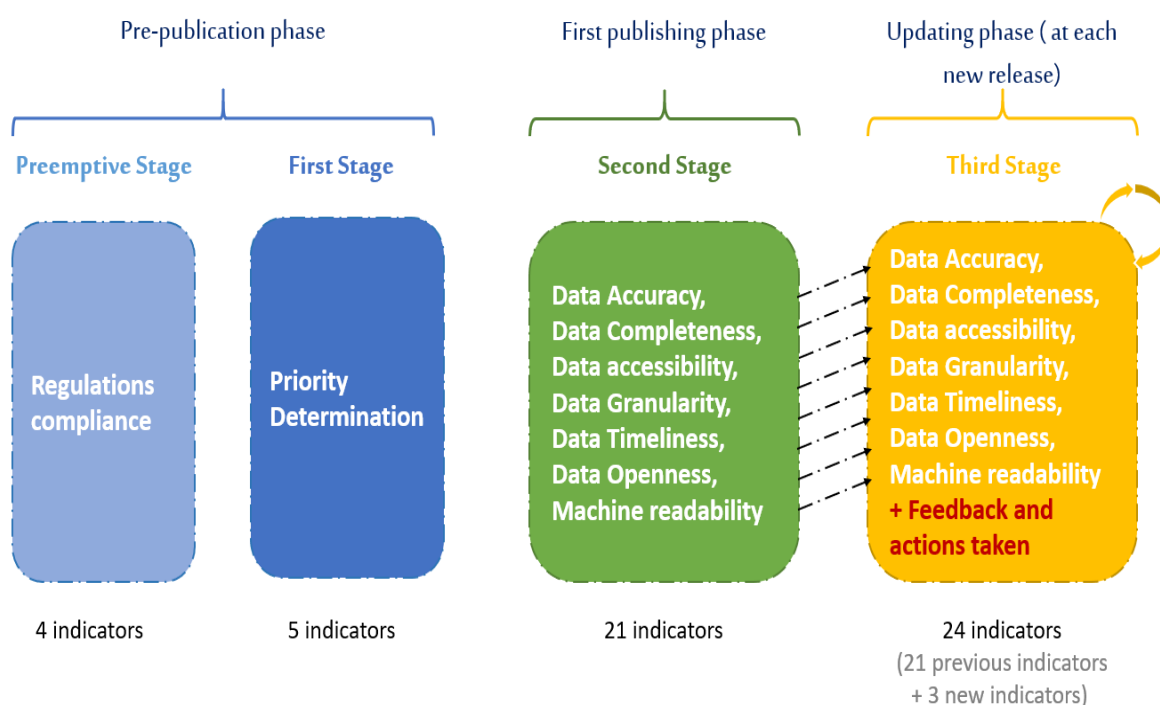
## The Mechanism of evaluating Indicators

### 1. Phases, dimensions, and indicators:

The procedural framework includes three main phases to assess and measure the quality of the open government dataset namely:

1. The pre-publication phase,
2. The first publication on the open data portal,
3. Updating phase (at each release).

Each phase corresponds to a set of dimensions and indicators, shown in the figure below:



### 2. Indicator Assessment Calculation:

The procedural framework includes a set of indicators; the government entity (specifically the data coordinator and portal content focal point) should audit these indicators. Two calculation methods are adopted depending on the indicator type, namely:

- 1.1. **Evaluation according to the achievement level (1.. 5):** which is based on measuring the level of progress in achieving the indicator. This method is applied whenever it is possible to define and measure the progress. The evaluation process is carried out by

## Quality Of Governmental Open Datasets

assigning the value of (1) to (5) according to the estimate of achievement, as explained within the table below:

Description	Score
If the indicator is not implemented or achieved at all	1
If the indicator is achieved with a percentage greater than zero and less or equal to 25%	2
If the indicator is achieved by a percentage greater than 25% and less or equal to 50%	3
If the indicator is achieved by a percentage greater than 50% and less or equal to 75%	4
If the indicator is achieved by a percentage greater than 75%	5

**1.2. Dual-options evaluation (1 or 5):** Some indicators may be evaluated with (Yes/No) answers by assigning the value (1) or (5) respectively:

Description	Score
If the answer to Indicator question is NO	1
If the answer to Indicator question is YES	5

### Calculating the score of Dataset Quality:

The final score at each stage is calculated as a percentage of the division of the sum of the obtained indicators values, over the number of indicators multiplied by 5.

$$\text{Quality Score (\%)} = \frac{\sum_{k=0}^n \text{indicators values (within specific stage)}}{\text{Number of indicators} \times 5} \times 100$$

## Quality Of Governmental Open Datasets

### Pre-emptive/Proactive Phase

### Regulations Compliance

At this stage, the data coordinator within the public entity will verify the degree of compliance of the data to be disclosed with the established legislative and regulatory systems.

To continue the evaluation through the next phases within this procedural framework, it is imperative that all the prerequisites mentioned in the table below are fulfilled.

→ It is highly recommended to share the results of this proactive phase.

Condition/ Prerequisite	Evaluation Responsibility	Level	Score Yes / No
Data meets the legal requirements of the 2007 access to information law No. (47) and its amendments upon issuance.	Data Coordinator	Only Data Level	5 or 1
Compliance with the requirements of the Government Data Classification and Management for the year 2020 and the instructions and guidelines issued therein.			5 or 1
The data set shall be free of personal data and comply with the provisions of the Personal Data Protection Act, upon issuance, and the regulations and guidelines published therein.			5 or 1
Adopting a license that allows sharing and the reuse of all published dataset.			5 or 1



## Quality Of Governmental Open Datasets

### The first phase

### Prioritization

At this stage, the data coordinator will define the priority level of opening up the dataset based on the government entity priorities and user's areas of interest (the closeness between data consumer need and data provider output).

No.			Indicator Title	Evaluation Responsibility	Level	Score
1	1	1	Promoting transparency principles	Data Coordinator	Data	1 .. 5
1	1	2	Empowering government performance and enhancing services and economic growth	Data Coordinator	Data	1 - 5
1	1	3	Supporting the strategic and operational goals within the government entity	Data Coordinator	Data	1 - 5
1	1	4	Internal demand for data (within the same entity or from others government entities)	Data Coordinator	Data	1 - 5
1	1	5	External demand for data (from the private sector, civil society, academics and individuals)	Data Coordinator	Data	1 - 5

### The first phase score value:

The quality score (as a percentage) for this stage, defined according to the following:

Quality Score	Priority level	Priority identification
Score < 65%	Low priority	The opening up dataset can be postponed until the higher priority data publishing is completed.
85% > Score => 65%	Medium priority	It is recommended to go to the next step (publication phase).
Total = >85%	High priority	Datasets with high priority, it should be published immediately.

## Quality Of Governmental Open Datasets

### The second phase

### Publishing Datasets first time

In this stage, the “**data coordinator**” and the “**data platform officer**” within the entity evaluate the quality of the dataset according to specific dimensions, namely: accuracy, completeness, granularity, timeliness, accessibility, machine readability and openness as shown below.

#### 1. Data accuracy:

Refers to whether the data values to be published reflect the data as it is stored in the business databases and to the degree with which data correctly represent the “real-life” objects they are intended to model. This dimension focuses on the evaluation of data values and their metadata and carried out by both “Data Coordinator” and “Data Platform Officer”.

No.			Indicator Title	Description of the indicator	Evaluation Responsibility	Level	Score
2	1	1	Data Quality audit system	The indicator rating is assigned based on the presence or absence of a data quality audit system, the results obtained, and the corrective actions taken by the entity (see explanation below)	Data Coordinator	Data	1..5
2	1	2	The ability of data to represent the “real-life”	This indicator focuses on the quality of the data to be published compared to the real-life (based on the principle of errors in the data compared to its real context).	Data Coordinator	Data	1..5
2	1	3	The ability of the data in its open format to represent the original data, i.e. as received from the data coordinator.	Ensure that the dataset matches the original data after its conversion to open format. The “Data Platform Officer” evaluate the data accuracy after converting the data into open format (the original data is provided by the “Data Coordinator”) (For example: checking that the converted file is encoded on UTF8, the conformity with CSV requirements (if the output format is CSV), when reshaping	Data platform officer	Data	1..5

## Quality Of Governmental Open Datasets

				data to vertical or horizontal representation, etc... <sup>1</sup>			
2	1	4	Metadata accuracy versus data content	Ensure that the metadata corresponds to the data to be published (the evaluation is based on the existence of a clear title and description, the relevance of the English / Arabic translation, the keywords and their adequacy with the content to be published).	Data platform officer	Meta Data	1..5

The indicator “2/1/1”: is evaluated by assigning values from 1 to 5, as below.

(1): If there is no system for auditing/processing the data to be opened.

(2): Existence of a system for auditing the data to be opened.

(3): Existence of an audit system for the data to be disclosed. The last audit dates back more than 5 years.

(4): The existence of an auditing system for the data to be disclosed. The last action is very recent (less than 5 years).

(5): The most recent data quality (internal or external) audit system assessed the data of very good quality.

### 2. Data completeness:

This dimension aims to assess the data requirements that must be clearly specified based on the data needs of the government entity and the data collection processes to meet these requirements (data have to be as complete as possible in order to reflect all saved/stored data without deficiency).

This dimension focuses on the evaluation of data values and their metadata and is carried out by both “Data Coordinator” and “Data Platform Officer”.

No.			Indicator Title	Description of the indicator	Evaluation Responsibility	Level	Score
2	2	1	Data Completeness and comprehensiveness	Data coordinator: is the person qualified to assess the completeness of the data. He will compare the data to be disclosed against the original content.	Data Coordinator	Data	1..5
2	2	2	There are no blank cells in the data file.	Depending on the number of cells with missing value (assign a value equal to 5, and subtract 1 at each missing cell appearance, until we reach the final score 1).	Data platform officer	Data	1..5
2	2	3	Availability of a data dictionary and its correctness	Depending on whether the data dictionary exists or not, with the extent to which it conforms to the	Data platform officer	Data	1..5

<sup>1</sup> Any use of encoding the data allows the preservation of the validity of the data, for example data in Arabic as in the case of UTF8 approved by Unicode

## Quality Of Governmental Open Datasets

				data content, especially after data conversion to an open format. A data dictionary includes a description and structure of each dataset field.			
2	2	4	Metadata completeness (mandatory indicator <sup>2</sup> )	This indicator evaluates the existence of required metadata (the title, the description, the key words, the translation to English/Arabic languages, etc.... ) according to five-stars scale	Data platform officer	Data	1..5
2	2	5	Data values for at least three periods date/time <sup>3</sup>	The tendency of the data to represent at least three consecutive time intervals (for example: if the data frequency is yearly, the dataset should contain values of at least three consecutive years 2018, 2019, 2020, etc....).	Data coordinator +Data platform officer	Data	1..5
2	2	6	Data linking within government entities	This indicator evaluates the potential of data linking with others Governmental agencies data owner.	Data coordinator +Data platform officer	Data	1..5
2	2	7	Data documentation	Data documentation will ensure that data will be understood and interpreted by any user. It will explain how data was created, collected, what the context is for the data, working assumptions, structure of the data and its contents.	Data platform officer	Data	1..5

### 3. Data accessibility & discoverability:

This dimension aims to measure the easiness to access to the open dataset in terms of URL naming pattern, Files that contain the data values, the metadata, etc...

It checks the data availability without any technical/legal/issue barriers, including whether the metadata is conforms to international standards. Moreover, it assesses the ability to obtain/download data (without human intervention).

<sup>2</sup> Positive indicators mean that no final result for the data quality scale is obtained without obtaining the presumptive value for this indicator, i.e. obtaining an evaluation equal to 5

<sup>3</sup> This indicator is calculated only for periodically released data sets

## Quality Of Governmental Open Datasets

No.			Indicator Title	Description of the indicator	Evaluation Responsibility	Level	Score
2	3	1	Automatically access metadata	This indicator assesses how metadata can be accessed and processed in an automated manner.	Data platform officer	Metadata	1 or 5
2	3	2	Metadata conformity with Dublin Core terms	The metadata is aligned to Dublin Core vocabulary at a minimum, and in conformity with the DCAT <sup>4</sup> standard. (Should be guaranteed by the open data platform)	Data platform officer	Metadata	1..5
2	3	3	Metadata conformity with Schema.org <sup>5</sup>	How easy it is to access the dataset through schema.org collection of terms. (Should be guaranteed by the open data platform)	Data platform officer	Metadata	1..5
2	3	4	Accessing the dataset via the open data portal's search engine.	The dataset visibility, namely the positioning of a dataset prominently in the results pages for certain keywords (internal search backend within the open data portal). It reflects the ability of the metadata (title, description, keywords in both Arabic and English languages) to match the requested dataset within the data portal.	Data platform officer	Metadata	1..5
2	3	5	Accessing the metadata through an API endpoint. (Application Programming Interface).	This will allow developers and other systems to access and process the metadata automatically utilising the API. (Should be guaranteed by the open data platform)	Data platform officer	Metadata	1..5

#### 4- Primary Data Source: (Data Granularity):

This dimension aims to ensure that open datasets should be in a form of basic raw materials as possible, the precise form in which the datasets have been collected from origin sources, with no aggregations or modification.

<sup>4</sup> This indicator is currently ineffective and will not be implemented for the time being until the new platform is completed.

<sup>5</sup> This indicator is currently ineffective and will not be implemented for the time being until the new platform is completed.

## Quality Of Governmental Open Datasets

No.			Indicator Title	Description of the indicator	Evaluation Responsibility	Level	Score
2	4	1	Level of detail compared to the original data	The extent to which the detail of the open dataset matches the data collected from the sources (granular and non-aggregated data considering some business constraints).	data coordinator + Data platform officer	Data + Metadata	1..5

### 5- Data timeliness:

Refers to the availability and accessibility of data on time. This concept is associated with the degree to which open data re-users have the data they need at the right time. Data re-users can be internal (within the Government entities) and/or external users. It helps to assess how the entity is disclosing open datasets on time according to the data requirements (without delays).

No.			Indicator Title	Description of the indicator	Evaluation Responsibility	Level	Score
2	5	1	Data dissemination time according to its requirements	That the data are published in a timely manner, and concurrent with the time of its collection and collection, taking into account the prioritization of data in which time is a critical factor in benefiting from it:  For example: budget data has to be disclosed once it is approved by the parliament	Data Coordinator + Data platform officer	Meta data	1..5

### 6- Data Openness:

This dimension assesses the datasets level of openness. The openness indicator can be determined based on the evaluation of the 5 stars table as shown below. The five-star scale is considered a cumulative rating, so that the level of the three stars includes the previous two levels.

1 star	Available on the web (whatever format) but with an open licence, to be Open Data
2 stars	Available as machine-readable structured data. (for example, Excel instead of scanning a table as an image or a PDF file).
3 stars	As (2 stars) plus non-proprietary format. (for example: CSV, XML, JSON, RDF instead of excel)
4 stars	All the above plus, use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff. The use of Unique Resource Identifier URI
5 stars	All the above, plus: Link your data to other producers data to provide context.

## Quality Of Governmental Open Datasets

No.			Indicator Title	Description of the indicator	Evaluation Responsibility	Level	Score
2	6	1	Technical openness level (TBL five-star)	According to the file format and its structure based on the five-star "TBL" model.	Data platform officer	Data	1..5

### 7- Automated processing:

This dimension targets the technical aspect of the data set by focusing on the ability to process the data content of the group by means of automated processes.

No.			Indicator Title	Description of the indicator	Evaluation Responsibility	Level	Score
2	7	1	The ability to process data automatically	The ability to process the data values automatically (without human intervention to improve its structure or content). For example, evaluate how much the data is conform to standards: ISO8601 for dates WGS84 for geographic data ISO 4217 for currencies ISO 639 for languages ISO 3166 for countries Etc....	Data platform officer	Data	1..5
2	7	2	Consuming the data using API endpoint.	According to the capability of consuming the data, using an API (application programming interface). The existence of an API endpoint or not.	Data platform officer	Data	1 or 5

### Calculations of the second phase score:

The quality score (percentage) for this stage is defined according to the table below:

Quality Score	Priority level	Explanation
Score < 65%	Low quality	The quality is very weak. It is necessary to review the entire process to improve the quality level.
85% > Score => 65%	Medium quality	The dataset can be published, and the quality score should be improved in the next update steps.
Score = > 85%	High quality	Maintain the same baseline of data quality and try to increase it gradually during the next updates.

## Quality Of Governmental Open Datasets

### The third phase

### Updating Phase – at each new release

After completing the previous phase and obtaining a sufficient score, the "Data Platform Officer" must publish the dataset on the platform and respect its update frequency as indicated in the metadata. New versions of the dataset should be released in accordance with the requirements of the third phase (update phase).

In this stage, the quality of the datasets is assessed based on the previous 21 indicators (the six (06) dimensions within the previous stage) in addition to three (03) new indicators. The three new indicators concern the previous **timeliness** dimension as well as the new dimension named "**re-users feedback and taken actions**" as explained below:

#### Data timeliness:

No.	Indicator Title	Description of the indicator	Evaluation Responsibility	Level	Score
3	1	Respecting the data update frequency	Data platform officer	Data	1..5
3	1	Respecting the metadata updating frequency	Data platform officer	Meta-Data	1..5

#### Feedback and taken actions:

No.	Indicator Title	Description of the indicator	Evaluation Responsibility	Level	Score
3	2	Corrective actions taken based on the feedback of the re-users	Data Coordinator	Data + Meta-Data	1..5

#### Calculation of third stage score:

The quality score (percentage) for this stage is defined according to the table below:

Quality Score	Priority level	Explanation
Score < 65%	Low quality	The quality is very weak. It is necessary to review the entire updating process to improve the quality level.
85% > Score => 65%	Medium quality	The dataset update can be published, and the quality level should be improved during the next updating time.
Score = >85%	High quality	Maintain the same baseline of data quality and try to increase it gradually during the next updates.