Open Government
Open Data Toolkit

VISION

ESCWA, an innovative catalyst for a stable, just and flourishing Arab region

MISSION

Committed to the 2030 Agenda, ESCWA's passionate team produces innovative knowledge, fosters regional consensus and delivers transformational policy advice. Together, we work for a sustainable future for all.

Economic and Social Commission for Western Asia

Open Government
Open Data Toolkit

United Nations
Beirut

Contents

1. Planning and Management of Open Data Programmes

A well-designed plan is essential for successfully implementing a sustainable open data programme with the desired impact. When executing open data programmes or projects, a one-size-fits-all approach will not be effective. Each country, city and individual governmental organization has its unique context.

This plan was developed based on global references that include a study carried out jointly by the United Nations Development Programme (UNDP) and the Partnership for Open Data (World Bank, Open Data Institute and Open Knowledge),[1] the European Data Portal's Open Data Goldbook for Data Managers and Data Holders[2] and successful international open data programmes such as those of the United Kingdom, Australia and other countries. The present toolkit is also based on the ESCWA Capacity Development material on Open Data.[3]

The plan consists of four phases, with each phase containing a number of key tasks.

#	Phase	Key tasks
1	Preliminary phase	• Run the baseline data maturity assessment; • Develop the open data strategy.
2	Development phase	• Develop the open data policy; • Create the open data portal; • Prioritize and publish the datasets that should be released first; • Recruit the central team for coordinating the open data initiative/programme; • Build human capacity.
3	Roll-out phase	• Publish more high-value datasets; • Improve organizational capacity; • Promote the use of the published open data.
4	Standard practice phase	Maintain the sustainability of the open data programme.

---

[1]   See: https://theodi.org/article/how-to-plan-and-budget-an-open-data-initiative.

[2]   See: https://www.europeandataportal.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf.

[3]   See: https://www.unescwa.org/publications/open-government-greater-public-sector-transparency-accountability.

A. Preliminary Phase

In this phase, the organization should aim to secure the political and administrative support required to successfully implement the plan and to develop the business case for the open data (OD) initiative.

This phase usually starts with a senior decision maker authorizing exploratory work and normally ends with a political declaration to launch the open data initiative.[4]

> **Example:** In Dubai, an open data committee was formed in December 2014 by decree of the city's crown prince and chairman of the executive council.[5] The committee initiated the Government's efforts in this domain, which later led to the establishment of the Dubai Data portal.[6]

1. Developing the Open Data Strategy

To develop an open data strategy, it is recommended to first consider the following:

- Make the process open and collaborative by communicating the strategy design plans to all stakeholders, including the public, and offering them continuous updates;
- Make it engaging by allowing the public to contribute to the design of the strategy through public consultation;
- Align the strategic destination with higher-level organizational or national objectives and agendas.

> **Example:** In June 2017, the Irish Government launched a public consultation on the draft of the country's open data strategy for the period 2017-2022.[7]

2. Outline of the Open Data Strategy

(a) Context and Background

This section should:

(1) Offer a review of the local environment in which this strategy was designed and will be implemented;
(2) Cover, at a minimum, other related plans and activities of the organization;
(3) Cover other political and socioeconomic circumstances in the city or country.

---

[4]   See: https://theodi.org/article/how-to-plan-and-budget-an-open-data-initiative/.

[5]   See: https://gulfnews.com/news/uae/government/dubai-open-data-committee-formed-1.1429600.

[6]   Dubaidata.ae.

[7]   See: https://data.gov.ie/blog/ministerdonohoelaunchespublicconsultationprocessforthenationalopendatastrategy2017-2022.

> **Example:** The data strategy for London (Data for London)[8] covers the context under the "Policy Context" and "Strategic Overview" sections. The two sections describe the data landscape in London prior to the strategy and the strategy's linkages to other policy areas, including energy policy and the broader smart cities agenda.

(b) Strategic Direction

This can be shaped as the "vision" or the "mission" of the programme, offering a clear description of the situation the programme aspires to achieve.

> **Example:** In Edmonton, Canada, the open data strategy has the vision of "leading the way to an open, smart and resilient city".[9] The keywords in this statement (such as resilient) should shape the remaining sections of the strategy, especially the objectives and principles.

(c) Objectives

The objectives provide a quantifiable version of the strategic direction. Consequently, they have to be specific, measurable, attainable, relevant and timely (SMART).

> **Example:** In Australia's "Open Data Strategy for the Department of Health in Queensland", objectives are grouped into themes. The theme of improving data quality has the following two objectives: ensure data is released within set standards and in accordance with legislative and other protections; and review currently available data sets to ensure links are active.[10]

(d) Stakeholders and Governance

(1) Identify the stakeholders of the open data programme (this may differ between Arab countries depending on their local political and socioeconomic contexts);

(2) Identify stakeholders' expected roles.

> **Example:** The Open Data Strategy in Ireland includes the following list of stakeholders: Political leadership, public sector organizations, legislatures, businesses and start-ups, civil society and citizens, information professionals and librarians, researchers and scientists and media. Additionally, the role of the legislature, for example, is to draft and enact any legislation needed at any stage of the OD programme.

(e) Actions

Identify the actions that need to be taken to achieve the defined objectives. As a whole, the actions should move from the current situation to the aspired one.

---

[8]   See: https://data.gov.uk/dataset/83215e9c-a55e-4f15-810c-8d3b59bce6bc/data-for-london-a-city-data-strategy.

[9]   See: https://www.edmonton.ca/city_government/documents/PDF/CityofEdmonton_OpenDataStrategy.pdf.

[10] See: https://www.health.qld.gov.au/__data/assets/pdf_file/0030/436809/open-data-strategy.pdf.

> **Example:** In London's data strategy, one action states that "we will assess, plan and build a pan-London Centre of Data Innovation".[11]

(f)   Metrics for Measuring Progress

In some organizations, these are called key performance indicators. Whatever the Gold book name, they should be carefully selected and frequently measured (for instance, quarterly or annually).

> **Example:** Three key measures in London's Data Strategy are institutional capability, social capability and innovation capability.[12]

(g)   Release of Open Data

At least one of the objectives should define the organization's target for its open data inventory and catalogue. This should explain the number of targeted datasets and their domains or departments/sections.

> **Example:** In the Open Data Plan of New York City, a detailed list of the datasets scheduled for release between 2015 and 2018 was identified, with the name of the agency that owns each dataset among other details.[13]

(h)   Principles

Principles are the set of commitments that guide the overall implementation of the strategy. This section enables the open data team to emphasize the key principles that are essential for promoting the culture of openness and participation.

(i)   Other elements

The strategy document may contain other sections, such as budget estimates and key success factors and themes under which the objectives and actions may be grouped.

---

[11]   See: https://data.gov.uk/dataset/83215e9c-a55e-4f15-810c-8d3b59bce6bc/data-for-london-a-city-data-strategy.

[12]   See: https://data.gov.uk/dataset/83215e9c-a55e-4f15-810c-8d3b59bce6bc/data-for-london-a-city-data-strategy.

[13]   See: https://www1.nyc.gov/assets/home/downloads/pdf/reports/2015/NYC-Open-Data-Plan-2015.pdf.

B. Development Phase

1. Developing the Open Data Policy

A well-written policy is an essential component for a mature and sustainable open data programme. The policy should address both internal and external stakeholders with the following aims:

(a) Outline the details of the organization's commitment to publishing, sharing and consuming data;

(b) Help internal stakeholders (such as data owners) identify and prioritize releases of open data according to a standard process;

(c) Help external stakeholders (such as developers) understand how the organization will be releasing its data and how they can be involved;

(d) Serve as a central reference for all stakeholders on how the various open data activities should be carried out and how the different stakeholders communicate and collaborate;

(e) Provide a common understanding of the essentials of open data (beginning with definitions) so that all stakeholders have the same expectations of the programme.

Similar to the strategy, the open data policy should be developed through an open and engaging process.

2. Open data policy outline

Open data policy outlines and full open data policies have been compiled in the following locations in the Arab region and beyond:

Jordan: The Open Data Policy released by the Ministry of Information and Communications Technology (national level).

Qatar: The Open Data Policy issued by the Ministry of Transport and Communications (national level).

United States: City of Seattle Open Data Policy (city level).

Australia: The Transport Open Data Policy released by Transport for New South Wales (department level).

The samples cover different countries and different levels (national, city and department levels) to demonstrate how some of the policy elements can be approached differently according to the governance level.

(a) Context and Background

At minimum, the context section should cover the following:

(1) Reference to the strategy or any other programme that led to the creation of the policy;

(2) Reference to any other legislation, policies or guidelines relevant to this policy or the practice of sharing Government data.

> **Example:** In Jordan, the open data policy of the Ministry of Information and Communications Technology refers to Jordan's Third National Plan 2016-2018, under the Open Government Partnership (OGP) Initiative, among other references.[14]

(b) Basics of Open Data

This section should, at minimum, cover the definitions of closed, shared and open data.

> **Example:** Qatar's open data policy dedicates a section to definitions and acronyms covering open data concepts. In addition, the policy explains the 5-Star Linked Open Data model.[15]

(c) Purpose and Scope

This section should include:

(1) The objective of the policy;

(2) Definition of the target Government organizations and NGOs expected to adhere to the policy (for instance, all Government departments in a city);

(3) Definition of the types of data targeted by the policy (for instance, all Government data owned by these organizations).

> **Examples:** The Open Data Policy of the City of Seattle states that the policy "covers data created, collected and maintained by the City of Seattle or by contractors or third parties on behalf of the City of Seattle".[16] Another example is Australia's Transport Open Data Policy, issued by Transport for New South Wales, which identified five objectives for the policy, including the simplification and facilitation of the release of appropriate data.[17]

(d) Policy Provisions or Directives

This is the core of the policy and should cover the organization's key messages on how to run the open data programme. Any good policy should cover the following elements:

Governance: Defines the various stakeholders' roles and responsibilities and the interrelations between them.

---

[14]   See: https://portal.jordan.gov.jo/OGD-Policy_en.pdf.

[15]   See: http://www.motc.gov.qa/en/documents/document/open-data-policy.

[16]   See: http://www.seattle.gov/Documents/Departments/SeattleGovPortals/CityServices/OpenDataPolicyV1.pdf.

[17]   See: https://www.transport.nsw.gov.au/system/files/media/documents/2017/open-data-policy.pdf.

Example: In Qatar, since the scope of the policy is national, it requires every Government agency to assign responsibility to a senior officer to oversee the administration and implementation of open data activities and the various responsibilities detailed in the policy. Moreover, the policy details the roles and responsibilities of the central agency (which is the Ministry of Transport and Communications) and other agencies in a separate appendix.

Data selection and release: Specifies the approach for identifying and prioritizing data for release, in other words, how the open data inventory and catalogue will be created and maintained.

Data licensing: Describes the licence under which datasets will be published, in addition to the process of clearing rights during data collection.

Example: The open data policies of Qatar, Jordan and Seattle require the use of an open licence and define that licence. However, the open data policy of Transport for New South Wales goes a step further and recommends Creative Commons as a standard licence to be used.

Privacy and security considerations: Defines what is considered personal data that can be used to identify an individual, and thus should not be included in published datasets. Open data programmes in the Arab States must give special attention to this issue considering that many countries still lack legislation regulating the issue.

Example: The open data policy developed by Transport for New South Wales reflects the legal framework concerning privacy and data security. It requires the removal of "personal or identifying information from datasets in line with the Government Information Public Access Act, the Privacy and Personal Information Protection Act 1998 (NSW) or other applicable legislation".

Data quality and publishing standards: Specifies the recommended standard formats, the use of metadata and other standards.

Examples: In Jordan, the policy addresses the issue at a broader level under the "Open Data Principles" section and requires publishers to create the needed processes to ensure adherence to the stated general principles. The policy in Qatar, however, provides more detailed measures and requires all datasets to be published in an open format, listing the formats that are considered open (CSV, Jason). The policy also defines metadata and requires it to be maintained.

Review and measuring progress: Identifies the metrics that will be used to measure the successful implementation of the policy; should be accompanied by a clear timeline and review process.

Example: The open data policy in Qatar contains a clear list of targets that all Government agencies need to accomplish and a clear timeline (in months) for each target. One target, for example, states that each Government agency is required to develop its own open data plan within five months.

Engaging data users: The users in this context are the external stakeholders (researchers, developers and the general public). This engagement improves the quality of published data, helps in guiding its release and ensures its ease of use.

> **Example:** The open data policy of the City of Seattle contains a section for stakeholder engagement with recommendations on how this engagement should be carried out.

(e)   Other Elements

The policy document may offer some standard managerial or technical templates or tools.

3.   Open Data Team

Most successful open data initiatives have a small central team, a task force that leads and coordinates the work within the organization and also across other involved Government organizations and NGOs if the programme is at the city or national levels. The table below lists the recommended job titles and their responsibilities. While viewing the roles suggested in the list, the following should be taken into consideration:

- The list is not exhaustive and merely provides a starting point;

- There is no fixed and one-to-one relationship between roles and team members;

- Certain roles may be more significant and crucial than others depending on the needs of the organization and the maturity of the open data programme;

- The central team is usually overseen by a steering committee or advisory board within the Government.

Role/Title	Suggested Responsibilities
Programme Sponsor	Secures political and legislative support and approvals as necessary.
Chief Data Officer	Leads the entire programme with its various tracks, including the development of the open data strategy, policy and data prioritization and release.
Legal Advisor	• In charge of the process of drafting the suitable open license; • Leads the development of any necessary legislation.
Portal/IT Manager	• Sponsors the development of the open data portal; • Provides advice and guidance on agency-specific technological issues.
Subject Matter Expert	• Provides detailed technical knowledge on open data; • Participates in regular review activities to ensure that datasets are published and updated as required.

Role/Title	Suggested Responsibilities
Coordinator	• Leads the communication work of the initiative within the Government, including briefing political sponsors and raising external awareness; • Leads communications with stakeholders and partners outside the Government, including the civic community; • Manages the process and participation in activities; • Arranges suitable training and capacity-building activities; • Acts as the programme's focal point for partners and stakeholders.
Data Publisher	Publishes data on the data portal as required by the approved process and quality standards.

Sources: https://www.gartner.com/smarterwithgartner/understanding-the-chief-data-officer-role/;
https://theodi.org/article/how-to-plan-and-budget-an-open-data-initiative/; and
https://data.sa.gov.au/sites/default/files/Toolkit/Open-Data-Process-Guide.pdf.

4. Capacity-Building

Training and capacity-building should continue throughout the entire lifespan of the open data programme. While certain initial open data training and orientation activities may be carried out during the preliminary phase, it is important to design a comprehensive and ongoing training programme for the open data team and public servants to drive the initiative forward and spread an open data culture.

The actual skill gap and training needs should be identified in the data maturity assessment conducted in the preliminary phase. Nevertheless, examples of the skills needed may include the following:[18]

• Basic concepts and practices of open data;
• Management the open data portal;
• Publishing of data and the use of the portal software by relevant staff;
• For data publishers, finding high-value datasets and prioritizing their release;
• Skills related to legal, privacy and security aspects;
• Skills needed for successfully engaging with the public and stakeholders.

The training programme may include study visits to countries and organizations with advancement programmes and initiatives in the open data domain.

---

[18]   See: https://theodi.org/article/how-to-plan-and-budget-an-open-data-initiative/.

C. Roll-out Phase

The main focus during this phase is on publishing increasing amounts of open datasets. This requires improving the internal capacity of the open data team and the organization(s) involved and promoting the use of the published data by the public and the target users.

Examples of the key areas organizations may need to focus on during this phase include the following:

(1) Processes surrounding the prioritization and release of open data need to be improved to cope with the increase in the number of published datasets and to respond properly to the needs and expectations of data users;

(2) The open data team will need to handle some difficult (or precedent-setting) cases of data release, including privacy and anonymization, among other issues;

(3) The open data portal may need to be improved to add features that were not available in its initial version. This decision should be taken in the light of the feedback from the open data team and users;

(4) Developer-oriented events such as hackathons may be held to promote the use of published datasets. Hackathons are design sprint-like events that gather both data suppliers and teams of application developers in order to generate novel uses of data;

(5) Data innovation challenges and competitions outside the developer community may be carried out as well, to promote the innovative use of data by different segments of society.

D. Standard Practice Phase

By this phase, the open data programme should have reached a stage where the publication and update of data are parts of the normal business and standard procedures of Government organizations. However, a smaller open data team or unit may still be needed to ensure that the policy and processes stay in sync with requirements and that the open data portal is working effectively to meet its objectives.[19]

The main focus should be on ensuring the long-term sustainability of the open data programme, and this is likely to require the following tasks:

(1) Frequent review and update of the open data standards and references (for instance, the policy);

(2) Update and maintenance of the open data portal;

(3) Continued publication of datasets in alignment with the prioritization plans and user needs;

(4) Frequent and continuous evaluation of the open data programme to assess the level of progress achieved and identify areas for improvement;

(5) Maintaining engagement with data users inside and outside the Government.

---

[19] Ibid.

2. Making Open Data Inventories and Catalogues

This section aims to help open data teams by providing guidelines for creating open data inventories and catalogues, which is an essential practice in any successful open data programme.

The terms "inventory" and "catalogue" are sometimes used interchangeably. They are, however, defined as follows for the purposes of this guide: An **open data inventory** of a Government organization is the master record of all datasets owned by that organization and eligible to be published as open data in a standard format. An **open data catalogue** contains datasets that have already been published as open data; in other words, it forms only a subset of the inventory.

Benefits of open data inventories:

(1) Enable the Government organization to set the boundaries of the high-value data that can be published as open data;

(2) Can be used to design a roadmap for publishing open datasets in alignment with the organization's priorities and resources;

(3) Can be used by the open data team to keep track of their progress in publishing potential datasets as open data.

Benefits of open data catalogues:

(1) Facilitate user access to published datasets, especially when they have features like search and tags;

(2) Enable open data teams to automate some data publication and update tasks.

The following four-phase process is recommended for creating open data inventories and catalogues:

Phase 1: Prepare a master list of all high-value datasets owned by the organization.

Phase 2: Classify each dataset in the master list to create an open data inventory.

Phase 3: Release the open datasets from the inventory to the open data catalogue.

Phase 4: Update both the open data inventory and catalogue.

A. Phase 1: Preparing a Master List of High-Value Datasets

High-value datasets are those that are more likely to make an impact once released as open data. The open data team at the organization should:

- Run brainstorming sessions to develop a master list of high-value datasets;

- This can be done by identifying the areas where data can be exploited to solve problems, improve efficiency and effectiveness and facilitate collaborations and by drawing a pathway for deciding what data should be published.[20]

The table below lists potential high-value datasets that are recommended by international open data practices.

Sector/category	Subsector/subcategory	Sample datasets
Education	Education providers	• Performance of education providers; • Locations of education providers; • Education providers' sanitation data.
	Public libraries	• Bibliographies of national libraries; • Library locations.
	Budgets and finance	• Government spending on education; • Budgets per locality or school.
Health care	Health-care providers	• Address and geocoded locations of registered health-care providers (e.g. doctor offices, clinics, hospitals, pharmacies). This dataset can also contain contact details; • Health-care services offered by each provider; • Quality and performance level of each provider based on relevant performance measures and opinion surveys.
	Disease, vaccination and prescription data	Data on the prevalence, risk and treatment of diseases and illnesses, by locality where possible.
	Budget, spending and contracting data	Data on public-sector finances, which may include public-sector spending for each sector, by locality and by public body (at an individual payment level) and budget, grant and contract data by department.
Transport	Public transport timetables	• Static and real-time timetables for various forms of public transportation (e.g. trains, buses, trams, ferries); • Prices for various forms of public transportation (e.g. trains, buses, trams, ferries).

---

[20] https://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf.

Sector/category	Subsector/subcategory	Sample datasets
	Public transport location data	• Geocoded location data for routes, stops and terminals for the various forms of public transport; • Accessible features available in public transport facilities (e.g. ramps and tactile tiles at platform edges).
	Traffic flow information	• Flow rates of traffic at specific locations; • Geocoded locations of traffic incidents.
	Budget, spending and contracting data	Data on public sector finances, which may include public-sector spending for each sector, by locality and by public body (at an individual payment level) and budget, grant and contract data by department.
Environment	Weather and pollution	• Real-time and historic data of temperatures, rainfall, wind speed and other observations; • Forecast data; • Data on various types of pollution including, but not limited to, a pollution inventory (regulated and otherwise) and pollutant concentration.
	Soil data	Data on the state of soil, such as location data of contaminated soil and expected soil conditions and nutrients.
	Cadastral map	A map showing the precise location (using GPS coordinates), dimensions, boundaries and ownership of land parcels, which may also include additional details such as unique identifying numbers and certificate of title numbers.
Sanitation and waste	Waste disposal sites	Geocoded locations of waste disposal sites.
	Toilet facilities	Geocoded locations of toilet facilities which may include the type of toilet facilities available where possible.
Governance	Elections	• Geocoded locations of polling stations; • Election results.
	Finance and contracts	• Transaction spending; • Tenders and contracts.
Economy and business	National and local accounts	• National budget (planned and spent); • Local budget (planned and spent); • Company/business register.
Crime and justice	Crime	Geocoded locations of crimes which may include the type of crime and an outcome (e.g. an arrest) where possible.

Sources: https://theodi.org/article/how-to-prioritise-open-data-to-drive-global-development/; and
https://www.diplomatie.gouv.fr/IMG/pdf/Open_Data_Charter_Annex_FINAL_13_June_2013_cle0ff8a3.pdf.

In addition to the datasets listed in the table, the central team may consider datasets that meet any of the following criteria:

(1) Datasets that are aligned with the country's national priorities and agenda;
(2) Datasets that are aligned with the organization's strategic objectives and priorities;
(3) Datasets that are in high demand by the general public.

B. Phase 2: Classifying Datasets in the Master List

In principle, Government data should all be open by default as possible. However, not all the datasets collected in phase 1 and listed in the master list are eligible for immediate release to the general public as open data. This can be due to any of the following reasons:

Legal: Releasing the dataset might violate an existing law or regulation.

Security: Releasing a dataset might threaten national security.

Privacy: A dataset may contain personal data that can be used to identify individuals once the dataset is released.

The open data team needs to run a classification exercise to check if any dataset in the master list should not be published as open data. The remaining datasets should be marked as open data and represent the open data inventory.

Data Spectrum

The data spectrum is a tool that can aid Government organizations in data classification and inventory building. Accordingly, the data spectrum model should be used to better understand and implement the data classification process.

The spectrum classifies data based on its legal status, so data can be 'closed' on the far left side, 'open' on the far-right side or somewhere in between.

Source: https://theodi.org/data-spectrum.
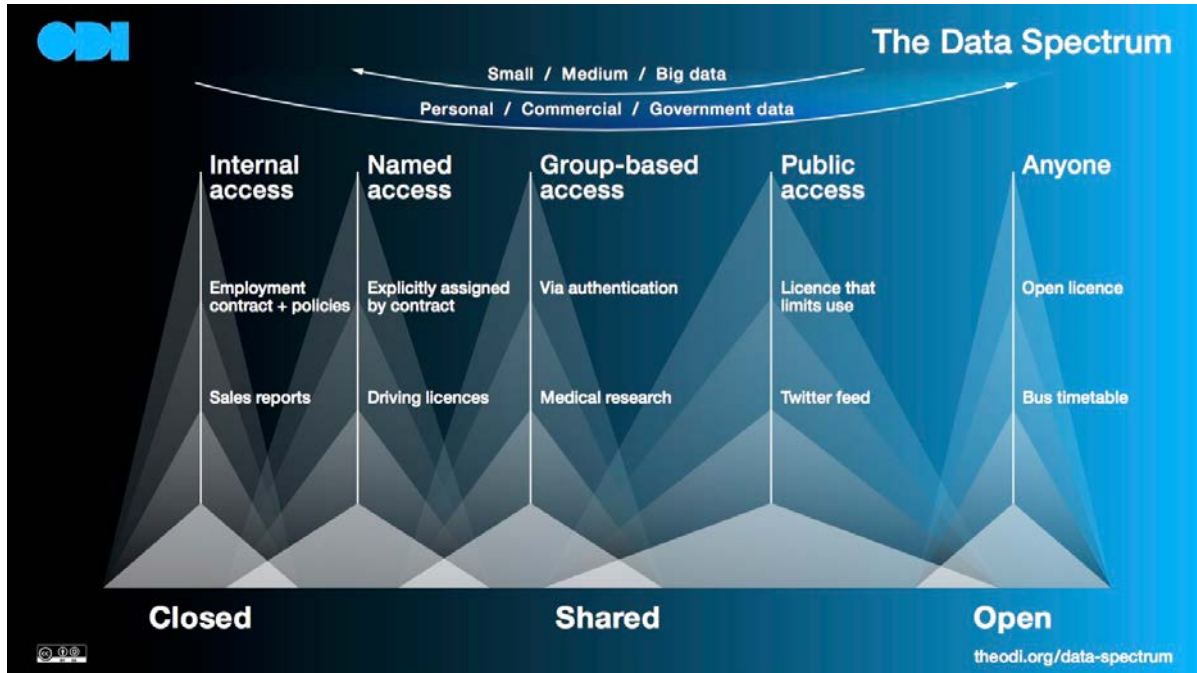
To elaborate further, below are brief definitions of the key classifications in the data spectrum.

Internal access: This classification, on the far left of the spectrum, relates to closed data that can only be accessed by some staff in the Government organization and is never to be shared outside it. The reason is that the data contains personal information.

> A good example are employee contracts, which usually contain personal information such as salaries, and thus should not be open to public access.

Named access: Moving from left to right on the spectrum, data begins to be shared outside the Government agency. Data labelled as "named access" can only be shared with named people or organizations outside the agency.

> An example of this are driver's licences, which contain personal data on drivers that they alone should have access to.

Group-based access: This is a more relaxed status where data is shared with a group of people outside the agency and not only with specific individuals, as is the case with named access. The group can be identified based on their job titles, the organizations they work for or any other factors that necessitate access to the data.

> An example of this are patient records in hospitals. Although these records contain personal information, they can be shared with insurance companies outside the hospital when needed or with a pre-approved group of practitioners in the medical field.

Public access: Data that is available to anyone under certain terms and conditions, so it is not yet open data.

> An example of this is a Twitter feed, which anyone can request access to for purposes such as the development of an application using those tweets, but Twitter will only grant access under certain limitations.

Anyone: To the far right is the data that can be described as open data. This is data that anyone can access, use and share.

> An example is the crimes dataset on the United States data portal.[21]

For an easier and more practical implementation of the data spectrum, Government agencies can just use the three key categories: closed, shared and open. **Closed data** in this categorization refers to data that is available exclusively to a particular individual, team or organization. **Shared data** refers to named, group-based or public access data. Finally, **open data** is that which anyone can access, use or share.

C. Phase 3: Releasing Datasets from the Inventory to the Open Data Catalogue

In this phase, the aim is to release datasets to the public by publishing them online on the open data catalogue. To do so, the following three processes should be carried out for each dataset:

Dataset selection

While all the datasets in the open data inventory should be of high value, the open data team may still need to prioritize the release of these datasets in alignment with the current demand from both the general public and the Government, potential of reuse, impact and according to the organization's internal capacity for dataset release and maintenance.

Dataset preparation

The open data team must at least take the following actions so that the dataset is ready to be published:

---

[21]   Data.Gov (2018). Crimes: 2001 - present. Available at https://catalog.data.gov/dataset/crimes-2001-to-present.

(a) Ensure the application of the appropriate open data licence;
(b) Use the appropriate open data format;
(c) Update the appropriate metadata;
(d) Check the dataset against the organization's standard quality criteria.

Dataset publication

After preparing the dataset, it is published on the online open data catalogue. This can be done in the following three different ways:

Manually: This is generally used for datasets that are relatively small and uncomplicated, and where the refresh rate is quarterly or longer.

Automated: In this method, software is used to automatically extract the dataset from where it is stored (for instance, a database) and publish it on the online catalogue.

Programmatic with application programming interface (API): APIs must be used to publish and update datasets with a live update frequency, namely, in real time or near-real time.

1. Open Data License

The data licence focuses on how permissions are granted to users for the use or extraction of the data published and owned by an agency or entity. Licenses may be either a public domain licence with no restrictions, which technically waives rights to content or data; an attribution licence, which explicitly states that users must attribute data used to the owner or publisher of the data; or an attribution and share-alike licence, which states that users must give attribution and share any derived content or data under the same licence.[22]

The following list shows the different ways a Government organization can acquire the open licence required for its open data:

Self-developed licence: a licence that is developed by the organization's own team for the sole use on its open data page or portal.

> **Example:** The Ministry of Interior in the **United Arab Emirates**[23] has developed its own licence for open data. The licence is presented on the open data policy page.

---

[22] Open Data Institute (n. d.). Publisher's Guide to Open Data Licensing. Available at https://theodi.org/article/publishers-guide-to-open-data-licensing/.

[23] https://www.moi.gov.ae/en/Open.Data/genericcontent/open.data.policy.aspx.

Government-wide licence: a licence that is developed by a central agency in the public sector and used by all agencies across the public sector.

> **Example: In the United Kingdom[24],** the National Archive developed the Open Government Licence as a tool "to enable information providers in the public sector to license the use and reuse of their information under a common open licence". Users are free to copy, publish and distribute the data, but they have to acknowledge the source of the information.

Creative Commons licence: The Government agency chooses and applies the suitable Creative Commons licence. Creative Commons is a global non-profit organization that facilitates the sharing and reuse of creativity and knowledge through its free legal tools, including a variety of licences.[25]

> **Example: The Australian Government[26]** follows a centralized approach but has not created its own licence. Instead, the Government uses Creative Commons. Users can share, copy, redistribute, adapt and remix materials, but they must provide appropriate attribution and indicate if any changes have been made.

It is strongly recommended to avoid the self-developed licence approach when possible and adopt one of the two other approaches instead. This is to avoid inconsistency across Government open data that could confuse users.

2. Dataset Format

The format of an open dataset refers to how the data is structured and made available to humans and machines and includes three elements: format, structure and delivery mechanism.[27]

Open data should be published in a format that enables anyone to access, use or share it. This means that the data should be non-proprietary and in a machine-readable format, regardless of what operating system or licensed software a user has access to. There is a wide range of file formats that can be used for publishing datasets, and the following table lists the most common of them.

---

[24]   http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/.

[25]   See: https://creativecommons.org/.

[26]   Creative Commons Australian Government License, (no date).

[27]   See: https://accelerate.theodi.org/.

Popular open data formats

No.	Format name	Definition	Type of data to use this for
1	Comma-separated values (CSV)	CSV is a great way to store large amounts of data with just commas separating the data values. CSV files often contain a header with names describing the data in the file.	Tabular data
2	Tab-separated values (TSV)	TSV is a very common form of text file format for sharing tabular data and is highly machine-readable.	Tabular data
3	JavaScript Object Notation (JSON)	JSON uses human-readable text to transmit data objects consisting of attribute-value pairs. It is used primarily to transmit data between a server and web application as an alternative to XML.	Complex-structured, multidimensional or tabular data
4	Extensible mark-up language (XML)	XML is a widely used mark-up language for encoding documents in a format that is readable to both humans and machines. Users create and define their own tags.	Complex-structured, multidimensional or tabular data, e.g. database extract metadata
5	Resource description framework (RDF)	RDF is a standard model for data interchange on the web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed.	Used for events/announcements
6	Geography mark-up language (GML)	GML is the XML grammar defined by the Open Geospatial Consortium for expressing geographical features. GML serves as a modelling language for geographic systems as well as an open interchange format for geographic transactions on the Internet.	Spatial/location data
7	GeoJSON	GeoJSON is an open standard format for encoding collections of simple geographical features along with their non-spatial attributes using JSON.	Spatial/location data

Source: https://data.sa.gov.au/sites/default/files/Toolkit/Open-Data-Process-Guide.pdf.

The structure of the dataset describes how the data is presented, with the most common structures being the following:

(a) Tabular data: data is presented in columns and rows;
(b) Hierarchical data: data is presented in a tree-like structure;
(c) Network data: data is presented in a flexible representation where nodes are not restricted to a hierarchy.

The delivery mechanism focuses on two key features, namely:

(a) The frequency of updating the datasets;
(b) Whether to make the dataset downloadable or not. If it is to be downloadable then its size should be small.

When selecting the format for a dataset, the open data team should consider the following:

(1) Choose the format that is most likely to be used by the general public;
(2) Publish the same dataset in different formats to accommodate different user needs;
(3) Publish a CSV version of datasets, as it is the most common format, in addition to any another format when needed.

3.   Metadata

Metadata is structured information that describes, explains, locates or otherwise makes it easier to retrieve, use or manage an information resource. Metadata is often called data about data or information about information.[28] Metadata is useful because it provides information enabling the user to make sense of the published dataset.

There are four general types of metadata:

Administrative metadata is the most common and is produced in data collection, production, publication and archiving. Most open data metadata is in this category.

Structural metadata describes the structure of a dataset, including its format, organization and variables. This is highest in demand by researchers and academics.

Reference/descriptive metadata is a broad term that mostly involves descriptions of methodology, sampling and quality.

Behavioural metadata records the reactions and behaviours of a dataset's users such as rating or user analytics.[29]

---

[28]   See: http://www.niso.org/publications/press/UnderstandingMetadata.pdf.

[29]   See: https://support.socrata.com/hc/en-us/articles/115008609707-Best-Practices-for-Metadata-Management.

> **Example:** The Police Force Strength data set of the Greater London Authority is published with a metadata set that includes the following:
>
> Name: Police Force Strength.
>
> Date published/created: 12 days ago.
>
> Licence: UK Open Government Licence.
>
> Format: CSV.

The table below lists the recommended standard basic metadata. It is useful to notice that the life cycle of metadata is longer than that of the data itself, which means that the metadata of a dataset may be created before the dataset itself and may be kept after the dataset has been removed.[30]

List of standard metadata

#	Metadata	Description
1	Title	Name of the dataset
2	Description	Free text
3	Keyword/tags	A keyword or tag describing the dataset
4	Publisher	Name of the Government organization responsible for releasing the dataset online
5	Themes	Business domains under which the dataset is classified (e.g. health care or transportation). A dataset can have multiple themes.
6	Language	Language of the dataset (e.g. English or Arabic)
7	License	Type of legal license the dataset is released under
8	Publish (or release) date	Date on which the dataset was first published
9	Update date	Most recent date on which the dataset was changed, updated or modified
10	Update frequency	How often the dataset is updated
11	Contact	Contact details of the publishing organization including, email and phone number if possible
12	Format	N/A

Sources: https://www.w3.org/TR/vocab-dcat/; and https://data.gov.au/dataset/data-gov-au-metadata-and-other-schemas/resource/c79c5839-008e-4833-b469-3c550d38d05c.

---

[30] See: https://www.seattle.gov/Documents/Departments/SeattleGovPortals/CityServices/OpenDataPlaybook_Published _2016.08.pdf.

D. Phase 4: Updating the Open Data Inventory and Catalogue

Once an open data inventory and catalogue have been created, the open data team needs to continuously maintain and update them. The time and effort needed for this depend on the size of both the inventory and the catalogue, in addition to the level of usage and engagement from the general public.

These continuous efforts include the following essential tasks:

(1) Adding more datasets to the open data inventory;

(2) Enriching the open data catalogue by selecting new datasets from the inventory and publishing them on the catalogue;

(3) Receiving and responding to any inquiries or requests from the public;

(4) Solving any issues, such as broken links, that may emerge;

(5) Updating the datasets as promised in the metadata.

3. Open Data Platforms

Publishing open data on the web makes it universally accessible, which satisfies an essential element of the definition of open data.

This makes the web the core channel for people to find and access open data. However, the web platforms used for publication can vary from one organization to another depending on their ambitions and their open data strategies.

In general, there are three main options Government organizations can choose from[31]:

Option 1: Online downloads.

Option 2: Data portals.

Option 3: Application Programming Interfaces (APIs).

A. Option 1: Online Download

This option is ideal early on when starting data publication, when the organization still has a limited number of datasets to publish. It is probably the simplest and least expensive way to publish open data. Datasets can be published on the organization's regular website by adding a dedicated page or section to the main website with links to the datasets.

A slightly more technically advanced alternative is to use the organization's content management system (CMS), such as Drupal and WordPress, to publish the datasets in combination with an internal database.

> **Example:** The Ministry of Energy and Industry in the United Arab Emirates follows this option in publishing its data (as shown in the figure). The data is published on a page on the Ministry's website where a list of datasets in various formats (such as PDF and XLS) that are labelled as open data can be found.[32] Clicking on any of these datasets will trigger the dataset to be downloaded.

---

[31]   See: https://www.europeandataportal.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf.

[32]   The open data page is a good example of direct online download. Available through: https://www.moei.gov.ae/en/open-data.aspx.

B. Option 2: Data Portal

This is a more advanced way of publishing datasets where the organization dedicates a whole portal to open data independent from its main website. This option is particularly useful if the publishing organization has large amounts of datasets to publish that need to be updated regularly.

To implement the data portal option, organizations can choose from two common scenarios:

Scenario 1: Create, own and run their own platform.

Scenario 2: Use another organization's platform. Usually, that other organization is the one in charge of the central national or city-wide open data programme.

In some cases, where the organization has a very strong digital presence and direct communication with its data users, it can adopt a combination of both scenarios.

> **Example:** In Tunisia, the central open data platform[33] harvests datasets from various sectoral platforms at national and local levels as well, based on DCAT standard. Please see the figure.

---

[33]   http://www.data.gov.tn/.

C. Option 3: Application Programming Interfaces (APIs)

APIs, rather than being treated as a stand-alone option, are widely adopted as an advanced feature on top of either of the previous two options. An API is an advanced way to connect distinct systems or applications to one another. In the context of open data, an API is a software that provides a system with direct access to data from another system.

The API grants third party applications direct access to the data catalogues and their functionalities via a specific request protocol, without mediation through any separate end-user interfaces. In short, it is particularly useful if the data must be up to date, directly accessed and reused by third parties, and the application using the data needs direct access to the database without any interference.

> **Example:** Transport for London has an API that enables developers to connect their software and access the data in a simple way. As a result, thousands of developers have used this API to build hundreds of mobile applications, including Citymapper, which is one of the most popular transport applications in the city.

D. Key Features of Open Data Platforms

The common features of open data platforms include the following:

Usability and accessibility: Portals should be designed and developed in alignment with international usability and accessibility standards such as those developed by the World Wide Web Consortium (W3C). Users should be able to access and use the datasets anonymously without user registration.

Data catalogue: When provided as a built-in feature, it facilitates the process of publishing managing datasets.

Metadata: A built-in support for metadata can enable the publisher to assign the required metadata to dataset(s) quickly and easily.

Data licence: Similarly, some data platforms enable the publisher to define one central licence and apply it to all or some datasets.

Dataset download option: The user should be able to easily download published datasets.

Data format: The portal should support a wide range of data formats (for instance, CSV, XML and JSON).

Data preview: This feature enables users to preview the data prior to downloading it.

Data visualization: This feature uses built-in graphing or mapping tools to enable the user to visualize the data.

Search function: The user should be able to easily search the data catalogue by a number of parameters such as subject, organization or data format.

> Example: On Data.gov.uk, users can browse data available within the catalogue by theme (for instance, education and environment), by publisher (for instance, Government agencies) or by format (for instance, CSV), among other options. In addition, the catalogue has a search engine with advanced features and result-filtering options.

Engagement: Open data portals should offer a space for dialogue between users and the data publisher. This space can offer some resources and user guides in addition to e-participation tools, such as discussion forums and social media communications, and it is important to include a form to request new datasets.

Application programming interface (API): As explained previously, APIs can enable software developers to build creative solutions using published datasets.

Security: Open data catalogues should implement security measures to protect data and metadata from being changed by unauthorized users.

Example: Data.gov provides details on the security measures used to protect the portal's data and content on its Privacy and Policies page. An example states that "commercially reasonable tools and techniques are used to protect against unauthorized access to Data.gov systems".[34].

Monitoring and analytics: This feature provides the data publisher and portal administrator with insights into the usage of the portal and the datasets.

E. Open Data Platform Solutions

The following provides an overview of the most common open data platform solutions used by governmental organizations and NGOs around the world to publish and manage open data. A brief summary is provided of each solution, its key features and whether it supports the Arabic language, in addition to three examples of open data portals. This list is not exhaustive, and it does not constitute any endorsement for any of these platforms.

The Comprehensive Knowledge Archive Network (CKAN): An open source for creating and managing open data websites. It can be used by national and local Governments, research institutions and other organizations that collect a lot of data.

Key features

Provides a wealth of features and has over 200 community extensions which can fill almost any feature gap.

Provides a streamlined way to make data discoverable and presentable. Each dataset is given its own page to list data resources and metadata, which adds to the value of the data catalogue and facilitates searching.

Supports Arabic language: Yes.

Examples of government open data portals that use CKAN:

https://www.data.gov/ (US).

https://data.gov.au/ (Australia).

https://www.europeandataportal.eu/ (EU).

http://www.agridata.tn (Tunisian Ministry of Agriculture).

---

[34]   See: https://www.data.gov/privacy-policy.

DKAN: A Drupal-based community-driven open source open data platform that offers a full suite of cataloguing, publishing and visualization features.

Key features

Community-driven feature development.

Built-in visualization features.

Manages diverse data sets.

Supports Arabic language: Yes.

Examples of government open data portals that use DKAN:
https://data.ca.gov/ (California, US).
http://data.gov.sa/ (Saudi Arabia).
http://data.gov.bd/ (Bangladesh).

OpenDataSoft: A turn-key solution fully accessible from a web browser that can be used by organizations to publish datasets and by users to visualize data in maps and graphs.

Key features

Built with an API-first approach, each dataset has an API with real-time data processing, querying and aggregation capabilities.

Intuitive data visualization tools.

Data processing and enrichment features.

Supports Arabic language: Yes.

Examples of government open data portals that use OpenDataSoft:
https://opendata.paris.fr/page/home/ (Paris, France).
http://spms.min-saude.pt/ (Ministry of Health, Portugal).
https://www.kapsarc.org/ (King Abdullah Petroleum Studies and Research Centre, Saudi Arabia).

Socrata: A cloud-based solution that allows Government organizations to publish their data online, make data-driven decisions, operate more efficiently and share insights with citizens.

Key features

Access a single source of trusted data in a self-service environment and break down silos and encourage collaboration.

Transform data into insights with visualizations and performance measurements.

Enhance the flow and quality of open data.

Increase transparency and resident engagement.

Supports Arabic Language: No.

Examples of government open data portals that use Socrata:

https://opendata.cityofnewyork.us/ (New York City, US).

https://data.cityofchicago.org/ (Chicago, US).

https://www.dallasopendata.com/ (Dallas, US).

4. Evaluation of Open Data Programmes

The following introduces a framework for evaluating open data programmes in Arab countries. The aim is to make evaluation as simple as possible, with clear steps and outputs. However, organisations should expect to go through a steep learning curve while building their human capacity and adjusting their organizational practices.

Before presenting the framework itself, it is important to highlight the following:

- Evaluation is not a stand-alone exercise isolated from the open data programme. Instead, it should be considered as an integral part of the programme;

- The evaluation process should be carried out continuously. This would provide the organization with experience in the process and also accumulate time-series data that can be very helpful in monitoring output trends of the evaluation process.

The following figure shows the framework, which consists of five steps covering the entire open data programme.
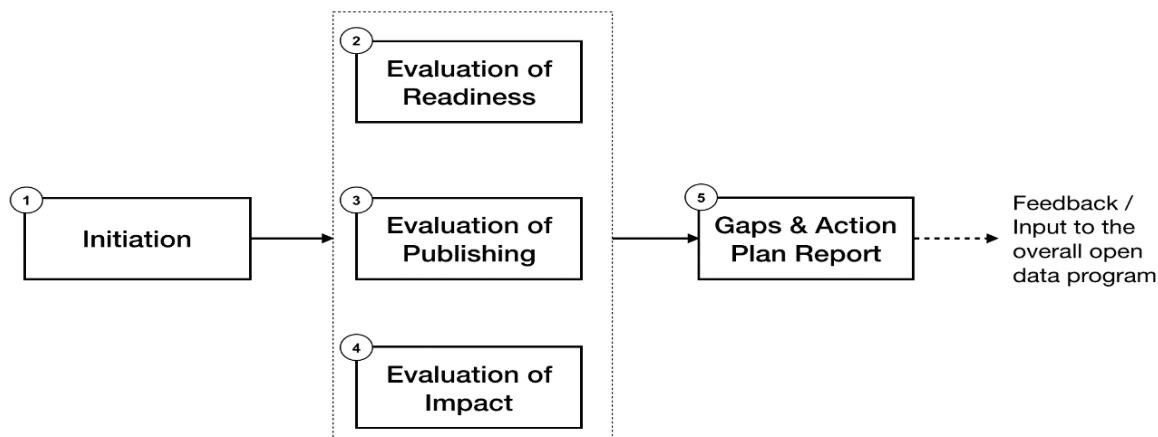
The core evaluation process occurs in steps 2, 3 and 4. The issues of readiness, publication and impact that are addressed in those steps are widely considered in similar open data evaluation models and methodologies, including those of the Open Data Barometer.

A. Step 1: Initiation

At the beginning of the evaluation, it is recommended to carry out the following steps:

(1) Secure the support of the organization's leadership, as this will increase the chances of conducting a smooth, honest and effective evaluation;

(2) Identify the evaluation team and team leader. Usually, this includes the team in charge of the open data programme, in addition to other stakeholders from inside and outside the organization such as representatives of data owners and external vendors who are in charge of some of the ongoing open data activities;

(3) Define the scope of the evaluation, which should encompass the entire open data programme. However, this scope might be too large for some organizations to handle all at once. In this case, they may decide to start by focusing on one part of this scope;

(4) Allocate the human and financial resources needed;

(5) Ensure that the evaluation process is well aligned with the open data strategy.

B. Step 2: Evaluation of Readiness

Objective: To evaluate the extent to which the organization has the required organizational and human capabilities to carry out a successful open data programme.

Output: A list of readiness gaps under the five key themes.

Methods and Tools: Open Data Maturity Model[35], Open Data Pathway[36] or Open Data Readiness Assessment tool (ODRA)[37].

In this guide, we will consider the Open Data Maturity Model covering the following five themes, which represent key areas of open data activities within the organization: Strategic Oversight, Data Management Processes, Knowledge and Skills, User Support and Engagement and Investment and Financial Performance. For each of these five themes, the model contains a set of organizational activities. There are 15 activities in total, shown in the table.

#	Themes of Open Data Maturity Model	Key organizational activities
1	Strategic Oversight	• Open data strategy; • Asset catalogue.
2	Investment and Financial Performance	• Financial oversight; • Dataset valuation; • Open data in procurement.

---

[35] Open Data Institute (2015). Open Data Maturity Model. Available at https://theodi.org/article/open-data-maturity-model-2/.

[36] Open Data Institute. See: http://pathway.theodi.org/.

[37] http://opendatatoolkit.worldbank.org/en/odra.html.

#	Themes of Open Data Maturity Model	Key organizational activities
3	Data Management Processes	• Data release; • Standards development and adoption; • Data governance; • Data desensitizing.
4	Knowledge and Skills	• Open data expertise; • Knowledge management.
5	User Support and Engagement	• Engagement; • Open data documentation; • Reuser support; • Community norms.

Process: To apply the Open Data Maturity Model, the organization should carry out an assessment exercise in which it reviews its current practices in each of the five themes. This exercise can be done internally (self-assessment) or with the help of external independent assessors. If the organization opts for self-assessment, it is recommended to use Open Data Pathway,[38] a free online assessment tool built to help organizations in assessments using the Open Data Maturity Model.

In either case, the following **eight-step** approach is recommended for a successful and smooth assessment process:

(1) Identify an organizational lead to coordinate the collection of information from across the organization and lead the entire assessment process;

(2) Identify the scope, which can cover the entire organization or focus on individual departments directly involved in the open data programme;

(3) Identify key participants. These are the people in the organization who may be needed to help answer specific questions or support the evaluation;

(4) Assess and score each activity (from the 15 in the table above) based on the result of the assessment. The following scale is recommended and defines five levels of achievement:

Level 1: Initial: The desired processes are non-existent or ad hoc, with no organizational oversight.

Level 2: Repeatable: Processes are becoming refined and repeatable, but only within the scope of individual teams or projects. There are no organizational standards.

Level 3: Defined: Processes are standardized within the organization based on best practices identified internally or from external sources. Knowledge and best practices start to be shared internally. However, the processes may still not be widely adopted.

Level 4: Managed: The organization has widely adopted standard processes and started to monitor them using defined metrics.

Level 5: Optimizing: The organization is attempting to optimize and refine its process to increase efficiency within the organization and, more widely, within its business sector.[39]

---

[38] Open Data Institute. See: http://pathway.theodi.org/.

[39] Ibid.

(5) Set appropriate targets for improvement after conducting a baseline assessment.

(6) Develop an action plan based on the results and the identified targets. The action items should be included in the plan, which is to be designed at a later stage of the evaluation;

(7) Circulate results, targets and action plans within the organization, including among those involved in supporting the assessment. Sharing this information more widely outside the organization can be considered as well;

(8) Set a date for the next assessment.[40]

By the end of evaluation of readiness, the organization should have identified a list of the activities that need to be improved, and should have also generated recommendations on how to improve them. At a later phase in the exercise, both the identified activities and recommendations will be incorporated into the gap report and action plan.

> The following presents an example to elaborate on the steps mentioned above:
>
> **Organization:** Department of Transport and Main Roads in Queensland, Australia.
>
> **Scope:** The entire open data programme: the department had 182 datasets published on its websites at the time of the assessment.
>
> **Lead:** The open data team, part of the information technology branch in the department.
>
> **Tools and:** The department ran a self-assessment exercise through a series of internal workshops attended by the open data team members and used the Open Data Pathway assessment tool. In addition, the department published all the results of this assessment on its website.[41]
>
> **Results:** The results of the assessment are summarized in the table below.

Theme	Score	Maximum Score
Data Management Processes	12	20
Knowledge and Skills	2	10
Customer Support and Engagement	6	20
Investment and Financial Performance	3	15
Strategic Oversight	4	10
Total	27	75

Source: Department of Transport and Main Roads, Australia, n. d.

---

[40] Ibid.

[41] Department of Transport and Main Roads, Australia (nd). Open Data Maturity Assessments. Available at https://data.qld.gov.au/dataset/department-of-transport-and-main-roads-open-data-maturity-assessments.

C. Step 3: Evaluation of Publishing

Objective: To evaluate to what extent the Government organization successfully publishes its datasets according to the predefined quality criteria.

Output: A list of gaps between the current status of publishing and the planned one.

Methods and Tools

The organization's own data inventory and catalogue plans (as discussed in section 2).

The organization's quality criteria (quality criteria will be discussed in detail in section 5).

The Open Data Barometer might be used in some cases.[42]

Process: Similar to the previous step, the evaluation of publishing can be carried out either internally via a self-assessment exercise, or with the help of an external independent assessor. The following steps are recommended for this assessment:

(1) Capture the current version of the open data catalogue that contains the latest list of open datasets published on the organization's open data page or portal;

(2) Use the list to evaluate dataset quantity and quality in the following manner:

- To evaluate the quantity of the published open datasets, the organization needs to simply compare the number of datasets published on the catalogue to the number of datasets it planned to publish under the open data strategy;

- To evaluate the quality of these published datasets, the organization needs to assess each individual dataset against the organization's quality criteria. For example, for each dataset, the format should be examined to confirm that it matches the list of formats approved by the organization as part of the quality criteria.

At the end of the process, the evaluation team should compile a list of gaps and issues and relevant notes about them, in addition to recommendations on how to fix them. These outputs should be incorporated in the gaps and action plan report to be produced at a later stage of the evolution exercise.

Example: The city of San Francisco runs such a quality and quantity evaluation of the datasets published by 52 government departments. The results of the evaluation are published on the city's data portal. This version of the evaluation shows that the city has published 437 datasets, representing 52 per cent of the inventoried datasets. The quality part of the evaluation covers the criteria identified by the city. For instance, the timeliness criterion indicates that 61 per cent of the published datasets are updated on time.

Source: https://datasf.org/opendata/.

---

42   World Wide Web Foundation.

D. Step 4: Evaluation of Impact

Objective: To evaluate the extent to which the open data programme successfully achieved its objectives as defined in the open data strategy.

Output: A review of the current gap between what has been achieved and the objectives initially stated in the open data plan, in addition to recommended action steps to bridge this gap.

Methods and Tools: A mix of quantitative and qualitative methods that the evaluation team identifies. Generally, the relevant literature refers to two main approaches to measuring the impact and benefits of open data:[43]

(1) Top-down (macroeconomic): This approach considers the value of open government data through the resources devoted to generating it or using it;

(2) Bottom-up (microeconomic): This approach seeks to find an aggregate figure by adding up various components using business surveys, local and international case studies and consultations.

> The following examples demonstrate how different levels of the impact of open data programmes are evaluated:
>
> The United States federal data portal (Data.gov) has a page dedicated to impact. The page presents a collection of case studies to demonstrate how the open data published on this portal has been used by start-ups and companies to create new businesses and jobs. For each company in the list, the actual datasets used by that company are identified, in addition to other important information such as the number of jobs created.[44]
>
> The EU has conducted a study to collect, assess and aggregate all economic evidence to forecast the benefits of open data for all 28 European member States and EFTA countries for the period 2016-2020. The study followed a mix of top-down and bottom-up approaches where the economic impact of open data was assessed using several tools including the market size of open data (in both euros and as a percentage of gross domestic product), the cost savings that can be achieved in the public sector, and the efficiency gains for individual citizens.[45]

E. Step 5: Gaps and Action Plan Report

This step represents the output of the evaluation team's work in the previous steps and should be allocated a minimum amount of time compared to the previous steps.

---

[43] Australian Government (2016). Open Government Data and Why It Matters. Available at https://www.communications.gov.au/file/14716/download?token=KSwxZ_B6.

[44] United States. For the page on impact, see: https://www.data.gov/impact/.

[45] European Commission (2015). Creating Value through Open Data. Available at https://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf.

Objective: To compile a report documenting all the issues captured in steps 2, 3 and 4 in addition to a set of recommended action steps on how the organization should handle these issues to improve the overall open data programme.

Output: A comprehensive document of all the gaps and recommended actions.

Methods and Tools: To create this report, the team should follow the standard policies and guidelines applied in the organization (such as communication guidelines). In addition, the team should consider publishing a version of the report online on the organization's website. The team can also use digital collaboration tools (such as Google Docs) to facilitate co-working among team members.

Steps: The following standard steps are suggested when creating the report. These steps can be amended depending on factors such as the organization's scope, team size and available resources:

(1) Appoint a moderator (or key author) for the report. The moderator's main job is to lead the work of drafting the report and facilitate the contribution of different team members;

(2) Agree on the overall structure (outline) of the report during the initiation phase. This should help in capturing all the information needed and minimize the risk of revisiting evaluation steps to recapture information. At a minimum, the report outline should include the following:

- Objective of the evaluation;
- Scope;
- Stakeholders inside and outside the organization;
- Evaluation of readiness: gaps and recommended corrective actions;
- Evaluation of publishing: gaps and recommended corrective actions;
- Evaluation of impact: gaps and recommended corrective actions;
- Action plan: full list of actions with targets and timeline;
- The planned date of the next evaluation.

(3) Capture all the issues and recommended actions throughout the evaluation phases. This should happen throughout the meetings and workshops conducted by the evaluation team;

(4) It may be helpful to use digital collaboration tools (such as Google Docs) to enable those involved in the evaluation exercise to jointly edit the report on an ongoing basis;

(5) Review the draft version of the report and obtain approval of the organization's leadership;

(6) Publish the report online. It is good practice to consider publishing the report (or parts of it) in different formats. For example, it may be useful to publish the evaluation of readiness section with the actual scoring as a spreadsheet to facilitate its access and use.

5. Open Data Quality

Publishing high-quality open datasets is crucial because open data of a poor quality can reduce user trust and negatively affect their use of the data. There are many different interpretations and ways of measuring data quality, including timeliness of publication, reliability and completeness.[46] Simply put, high-quality open data is data that is usable, and open data becomes usable "when a human can understand it and a machine can manipulate it".[47]

It is worth mentioning that data is used by machines (such as mobile applications) and this issue can become critical for Governments and cities that have smart-city and artificial Intelligence initiatives, examples of which are the United Arab Emirates, in particular Dubai, and should therefore be given special attention.

Defining an exhaustive list of detailed criteria for open data quality can be challenging considering that quality can be interpreted in different ways. However, there is some degree of consensus on the outlines of open data quality. The European Commission[48] has compiled the following criteria that closely align with the principles of the Open Data Charter and that also overlap with the quality parameters highlighted in the United Nations e-Government Survey 2016.[49] These criteria are as follows:

(a) Accuracy: Is the data correctly representing the real-world entity or event?
(b) Consistency: Is the data free from contradictions?
(c) Availability: Can the data be accessed now and over time?
(d) Completeness: Does the data include all the items that represent the entity or the event?
(e) Conformance: Is the data following accepted standards?
(f)   Credibility: Is the data based on trustworthy sources?
(g) Processability: Is the data machine-readable?
(h) Relevance: Is the amount of data included appropriate?
(i)   Timeliness: Does the data represent the actual situation and is it published soon enough?

As an easier alternative for translating the dimensions of open data quality into a specific set of applicable quality criteria, the Open Data Certificate may be used. Open Data Certificates were

---

[46] Open Knowledge International. Open Data Quality: the Next Shift in Open Data? Available at https://blog.okfn.org/2017/05/31/open-data-quality-the-next-shift-in-open-data/.

[47] European Union Data Portal.

[48] See: https://joinup.ec.europa.eu/sites/default/files/document/2015-05/d2.1.2_training_module_2.2_open_data_quality_v1.00_en.pdf.

[49] United Nations (2016). United Nations e-Government Survey. Available at https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2016.

developed by the Open Data Institute and are used or endorsed by governmental organizations and NGOs, including the European Commission.[50] The certificates are used to assess and recognize the sustainable publication of quality open data and are based on a set of criteria that are grouped under four categories: **legal, practical, technical and social**.[51]

A. Legal Criteria

The legal criteria cover issues related to rights, licensing and privacy and are described in the table below.

Criteria	Description
Open license	The statement that describes people's right to use the dataset and explains what they can and cannot do with the data.
Clear rights statement, detailing any copyrights	Needed if the data or part of it was created by the organization's intellectual effort, for example, by writing text that is within the data, or deciding whether particular data is to be included. There is no copyright if the data only contains facts where no judgements were made about whether to include them or not.
Privacy issues	How does the organization protect people's privacy in the published data? Can individuals be identified from this data? If there is such a risk, then the dataset needs to be de-identified before being published.
Machine-readable rights statement	It is good practice to embed information about rights in machine-readable formats so people can automatically attribute this data back to the author when they use it.

Source: Open Data Certificate (2017). Available at https://certificates.theodi.org/en/datasets/215670/certificate.

B. Practical Criteria

The practical criteria ensure that open data can be found and relied upon by users.

Criteria	Description
Accessibility	Data is available on the Internet for the widest range of users, including people with disabilities. The data should be in a format that enables its reuse.
Discoverability	Data can be found more easily if, for example, it is linked to the organization's main website or to its publications, or if the dataset is listed somewhere else (e.g. as part of a collection of related datasets).

---

[50]   See: https://www.europeandataportal.eu/elearning/en/module5/#/id/co-01.

[51]   Open Data Certificate (2017). Available at https://certificates.theodi.org/en/datasets/215670/certificate.

Criteria	Description
Data is time-stamped	The date of publishing and update should be part of the metadata associated with the dataset. This ensures that people using the data do not unintentionally use out-of-date information.
Data availability	Will the dataset remain published for a long time? Or could it disappear at any time?
Data timeliness	Data should always be up to date.
Data backup	Backups can be performed automatically depending on the technology used by the organization. In all cases, people need to know that they can rely on the availability of the data.
Quality documentation	Each dataset should have a uniform resource locator (URL) (can be part of the metadata) where there is documentation of the quality of data and any issues with that quality, so as to clarify the reliability of the data.

Source: Open Data Certificate (2017). Available at https://certificates.theodi.org/en/datasets/215670/certificate.

C. Technical Criteria

These criteria cover technical issues, such as the format in which the data is published.

Criteria	Description
Machine-readable format	Structured format that can be automatically read and processed by a computer, such as CSV, JSON, XML. PDF is not a machine-readable format.
Open standard machine-readable formats	Open standards enable anyone to implement them and are widely supported, so it is easier to share data with more people. XML, CSV and JSON are all open standards.
Machine-readable provenance documentation	Clarifies the origins of how the data was created and processed before it was published. It builds trust in the data because it enables people to trace back how it has been handled.
Content-appropriate formats	Different data formats suit different datasets depending on their contents. For example, geographical data can be plotted on a map as points or boundaries, whereas a tabular format can be more suitable for statistical or numerical data (e.g. census results).
Single consistent URL	A URL for accessing and downloading the data.
URLs as identifiers	Data is usually about real things, such as schools or hospitals, or uses a coding scheme. If data from different sources uses the same persistent and unique identifier to refer to the same things, people can combine sources easily to create more useful data.

Source: Open Data Certificate (2017). Available at https://certificates.theodi.org/en/datasets/215670/certificate.

D. Social Criteria

The social criteria aim to make the datasets more usable by human consumers.

Criteria	Description
Data is documented	Documentation about the data helps users understand the data's context and content. For example, documentation about the data API can help developers understand and use it.
Contact details provided	Providing a link (URL) to the webpage containing the contact details of the person(s) responsible for the data or who can answer questions about it.
Machine-readable metadata (documentation)	Enables users to easily access and use the metadata.
Social media accounts	Social media can be used to promote the data and connect with its users. If the organization uses social media, the accounts should be provided as part of the contact details.
Discussion page or forum	This could simply be the webpage for the dataset itself where people can post a comment to have a discussion, or a dedicated forum on the data portal or the publisher's website.
Dedicated communication team	This team will engage through social media and blogging and arrange "hackathons" or competitions to encourage people to use the data.

Source: Open Data Certificate (2017). Available at https://certificates.theodi.org/en/datasets/215670/certificate.

While applying all four categories of quality criteria, the open data team will likely notice that the level of implementation of most of these criteria varies from one organization to another, depending on the organization's needs and capacities. For example, one organization may not have the required human capacity or financial resources to frequently update a dataset.

Therefore, it is recommended that Government organizations design their own version of the quality criteria, based on open data principles, but tailored to the organization's business domain, level of readiness, user expectations and other factors particular to the organization. The organization must observe a high level of commitment to quality and must also continuously improve them over time to boost the quality of their data.

E. Open Data Quality Assurance Process

To help data publishers in the Arab region adopt and develop their own customized versions of the quality criteria presented previously, it is recommended to use the following standard process that

can facilitate their application. More about this subject could be read from the ESCWA Capacity Development Material on Open Data.[52]

The Quality Assurance Process

The process consists of the following six steps, namely: Define a standard set of criteria; Share with stakeholders; Inform and train stakeholders; Apply to new datasets; Check the quality of published datasets; and Review and update the quality criteria.
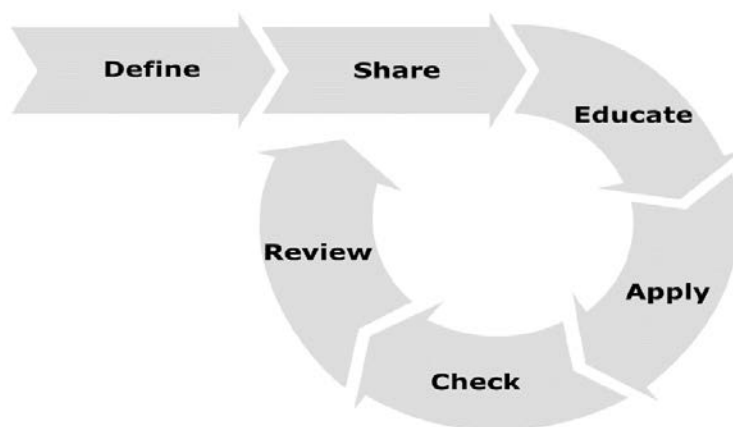
(a)   Define the criteria

The open data team should create a standard set of criteria approved by the organization. These criteria will be applicable to all datasets published under the open data programme. This process can be completed as part of the open data policy design process.

Recommended steps to that end may include the following:

(1)   Studying the list of four categories of criteria discussed in the previous sections;

(2)   Identifying the quality needs of the organization as determined by its business domain or the range of its datasets, in addition to users' expectations of quality. These needs and expectations should then be matched with the appropriate level of quality;

> For example, if the organization's core business produces a lot of Geospatial Information System (GIS) data, then it is essential to use open machine-readable geographic data formats.

---

(3) As recommended in many parts of this guide, the process of defining the quality criteria should be open and collaborative.

(b) Share the criteria with stakeholders

Once the quality criteria have been defined and approved, the team should share them with all stakeholders, including the public, to promote their usage. The quality criteria can be shared through several channels.

(c) Inform and train stakeholders

Similarly, the open data team should offer all stakeholders the training and information needed to explain the process and prepare them to follow it and make use of its outcomes. It is good practice to offer these training activities via both online and offline channels to accommodate the needs of different stakeholders.

(d) Apply to datasets

Every dataset has to pass a quality check before being published. This quality check will rely on the quality criteria designed in this process.

(e) Check datasets for quality

The open data team should carry out a periodic and ongoing quality control procedure to review published datasets and fix any possible quality issues.

(f) Review the quality criteria

The quality criteria themselves need to be reviewed and periodically updated. This may take place once a year. The feedback and suggestions for changes to the criteria can be collected from the following two main sources:

Internally: from the open data team and internal stakeholders. The team should look for suggestions for improving the quality criteria to reflect the development in the organization's readiness (for instance, human or technological) or changes to the datasets (for instance, need to publish live data).

Externally: from all segments of data users (for instance, researchers or software developers) who may have new requests (for instance, to offer an API for certain datasets). Every time a new version of the quality criteria is developed, the team should go through the process of sharing and educating.