



Data prioritization, preparation and publication

Division for Public Administration and Development
Management (DPADM), United Nations Department
of Economic and Social Affairs (UNDESA)

Content

Step-by-step guide, based on best practice, to identify, prioritize and prepare for the publication of key datasets.

The following guide consists of thirteen essential steps to identify, prioritize, prepare and publish OGD. Each of the steps provides the basic information and recommendations needed to move forward. The step-by-step guide is designed for government officials in charge at the level of data holding agencies.

13 steps to publish OGD

1. Start small and keep it simple
2. Understand supply and demand
3. Choose and prioritize datasets
4. Protect privacy
5. Add context
6. Structure data
7. Select an open license
8. Publish data online
9. Ensure data accessibility and availability
10. Raise users' awareness about agency's data and listen to their feedback
11. Foster data reuse
12. Improve data skills
13. Keep data fresh and preserved over time

Before starting to plan for an OGD initiative

Before starting to plan for an OGD initiative, it is recommended to think about opening data as a process that leads to change and can create social, environmental and economic value within society. Publishing government data is not just about technology and data.

To be useful, OGD needs to be used. So this process needs to start with a reflection on the potential users of data and the ways data might be made useful for others. The reflection should also include possible intermediaries that can help turn data into useful analysis, visualizations or tools.

With this in mind, the following steps can be followed as a simple guide to OGD.

1. Start small and keep it simple

Numerous agencies have started their OGD initiative with only a few datasets. Limiting data volume helps thinking and acting in an agile manner.

It is critical to:

- Get early feedback on shared data;
- Create awareness of the new platform among co-workers;
- Learn from feedback and change the approach as needed.

‘Keep things simple’. There is no need to try to do everything at the same time. A step-by-step and iterative approach should be followed to enable learning and improving over time. Technology should not be a major concern; there are several tools and platforms that help streamline this process.

2. Understand supply and demand

Data being published needs to be clear and user-friendly.

To thoroughly understand supply and demand it is critical to:

- Identify and invite potential users of the agency's data;
- Listen to their demands;
- Identify potential gaps between supply and demand;
- Inform future data collection and publication accordingly.

This is not a one-time exercise; it should rather be done in an iterative way.

3. Choose and prioritize datasets

Starting by publishing all data at the same time is not easily doable. Thus, it is recommended to prioritize the datasets the agency intends to publish first. Priority should be given to datasets that have the best balance between being:

- Available for publishing (the low hanging fruits);
- Datasets of high-quality;
- Datasets for which there is high demand;
- Useful to meet specific national priorities, including those related to the SDGs.

Having this balance increases the chance that data will be reused immediately after its publication. The positive results that this approach generates, will help mobilize support for the process of opening more datasets.

4. Protect privacy

Privacy is an internationally recognized human right. The right to privacy encompasses the right to protection of personal data. Some government datasets (e.g statistical data) are based on individual data about the population. It is critical to use anonymization techniques to make sure that no personal data appears in the dataset before publication.

It is recommended to make sure that:

- The agency's open dataset does not contain personal data;
- If the dataset has been anonymized, it cannot be de-anonymized.

There are established techniques on data anonymization. However, there are also methods to de-anonymize data. It is important to ensure that the method used for anonymization is effective against de-anonymization. In case of uncertainty on whether the dataset meets the standards of privacy and data protection, it is recommended to seek expert advice. Usually, the National Statistical Office or relevant agency should be able to provide advice on this issue.

5. Add context

Context will give re-users a better understanding of OGD. This context is usually contained in metadata. These include:

- Title;
- Description;
- Keywords;
- Legal information;
- License and attribution required;
- Geographical and temporal coverage;
- Granularity and completeness;
- Date of publication, when data was last modified and updating frequency requirement.

This information will help users better understand the content of data, how to interpret it and if it meets the user's specific needs. Metadata can also contain references to more detailed documentation, user manuals, or methodologies.

6. Structure data

To be used and reused, OGD should be available in well-structured and standard formats. To this end, data should be exported in one or several of the following structured formats:

- XML;
- XSL;
- CSV;
- JSON and GeoJSON;
- RDF.

This way data can be processed automatically by machines (sometimes referred to as machine readability). Unstructured information such as handwritten documents, or formats like Microsoft Word, PDF, or scanned images is not considered open data.

7. Select an open license

Explicit licensing is mandatory when publishing OGD so it can be reused by anyone for any purpose, including commercial reuse, without any restrictions. It is recommended to use a well-established license for Open Data to maximize interoperability. Examples include:

- Public Domain Dedication;
- Creative Commons Zero;
- Creative Commons Attribution 4.0;
- Creative Commons Attribution-Sharealike 4.0.

The latest version of Creative Commons license can be used as a template.

It is accepted practice to ask for attribution, but it is advisable to keep this as minimal as possible, e.g. name of the agency and a link to the license text.

8. Publish data online

OGD is truly open when it is easily retrievable and accessible. To start with, it is fine making data available for download on a simple website or on a FTP server. However, this does not automatically mean that the data can be easily found. Improving access to data requires to:

- Make it available online or set up a portal;
- Have access to it for free and anonymously (e.g. through a non-mandatory sign-up to a website/portal).

To increase the accessibility of a public agency's data, it is recommended to publish such data into an Open Data portal. This portal will allow gathering data in one single place and easily generating visualizations and Application Programming Interfaces (APIs). End-users will also be able to readily browse through datasets and use an advanced search for specific data.

9. Ensure data accessibility and availability

Using consistently Uniform Resource Identifier (URI) for any published dataset ensures that the data can be accessed under the same Uniform Resource Locator (URL) on the Internet over time. It is also advisable to make datasets available for bulk download (for example as an export of an entire database). A way to strengthen data availability is through generating APIs. In this regard it is important to:

- Use persistent URIs as identifiers for datasets downloads;
- Make data available as bulk download;
- Make data available via well documented APIs.

APIs allow re-users to build applications around data. Thanks to APIs, every time the originating agency updates its dataset, it will synchronize up with applications from third parties. APIs are especially useful for making real-time data accessible. APIs are great but should not replace bulk data download.

10. Raise users' awareness and listen to feedback

Once the data is made available online, the data holding agency needs to make sure that the public is informed about it. This requires to:

- Tell “the world” about published data
- Ask feedback for future improvements

It is important that the agency receives feedback from data users and uses this information for continuously improving data. Recommended modalities for engaging with potential reusers include social media or follow-up roundtables.

11. Foster data reuse

It is important to remember that data is published so it can be reused to create value. The agency needs to engage with potential reusers and intermediaries to help them make data useful for others, including vulnerable segments of the population, who may not otherwise benefit from this information. To this end, it is recommended to invest time and resources in dedicated programmes that foster reuse uptake. This may include to:

- Promote the reuse of data via hackathons, workshops or conferences;
- Engage with new intermediaries including those from the private sector and civil society, to expand data outreach and enhance its usefulness;
- Foster reuse by dedicated programmes like mini-grants and fellowship programmes, or business incubators, accelerators.

It is advisable to go beyond hackathons and partnering with others to set up dedicated programmes. These, in turn, support entrepreneurs and startups to engage with agency's data to develop new services and applications. Every successful business that has been built on data published by an agency is a success for the agency as well.

12. Improve data skills

Data can be useful not only to the public, but also (and mainly) to the agency which publishes it as well as other government agencies. Publishing data following open data principles makes it easier to access and reuse it. However, government officials need adequate technical and data analytical skills to effectively use OGD. To this end, a government agency needs to:

- Introduce capacity development programmes to leverage its officials' technical and data analytical skills;
- Help other government agencies to understand and use its data.

Because capacities on data literacy, data management and data analysis need to be enhanced within many different types of organizations both inside and outside government, it is highly recommended to explore potential collaborations among OGD national stakeholders around capacity development.

13. Keep data fresh and preserved over time

Outdated datasets often become useless. So it is critical to update data whenever possible and publish it under a dedicated new version that clearly differentiates with the previous ones. This does not mean “unpublish”/remove access to older versions. In some cases, it is essential that historical data remains available for comparison. So it is highly recommended to ensure long term preservation and availability of data over time. In sum, it is important to:

- Keep data fresh;
- Use version control for updating data;
- Ensure long term preservation and availability over time.

OGD is a long-term contribution to the community it serves. This requires long term commitment, it takes time, human and financial resources and energy but it won't be long before the agency can benefit from its OGD and leverage it to enhance its responsiveness to the public it serves.

Repeat the Sequence

Once an agency has successfully gone through the essential steps of identifying, prioritizing and publishing some datasets in open formats, OGD can be promoted and reused.

The agency should analyze feedback from potential users to make adjustments and improvements. Agencies can also receive feedback from the potential reusers about existing demand for other datasets. This provides input for the next round of datasets to be identified, prioritized and published as OGD.

Thank you!

Division for Public Administration and Development Management (DPADM) of the United Nations Department of Economic and Social Affairs (UNDESA).

OGD Project and OGDCE Guidelines
publicadministration.un.org/en/ogd

A United Nations Publication. Copyright © United Nations, 2017. All rights reserved.