Data prioritization, preparation and publication

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

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

Opening data is a process that leads to change and can create social and economic value within society. Publishing open data is not just about technology. Open data needs to be used to be useful.

• CONSIDER the potential users of your data and the ways your data might be useful for others.
• Think about possible intermediaries that can help turning your data into useful analysis, visualisations or tools.

With this in mind, you can follow the practical steps suggested as a simple guide to open data.
1. Start small and keep it simple

Numerous organizations have started their Open Data program with only a few datasets. Limiting data volume helps you think and act agile.

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

Keep things simple. Don’t try to do everything at the same time.

Follow a step-by-step and iterative approach, that enables learning and improving over time.

Don’t worry too much about technology, there are several tools and platforms to streamline the process.
2. Understand supply & demand

Understand the demand side - actual or latent users of your data.

- Identify and invite potential users of your Agency’s data
- Listen to their demands. Invite them to roundtables to discuss with those responsible for supply of data in your agency.
- Compare supply and demand and identify potential gaps between supply and demand: This gap-analysis will help you create specific recommendations for future data collection and publication that is useful.
- Inform future data collection and publication accordingly

This is not a one-time exercise, but should be done in an iterative way. It is recommended actively collecting feedback and using it for further improvement.
3. Choose and prioritize datasets

Prioritize datasets

- You cannot start by publishing all data at the same time.
- Priority should be given to datasets that have the best balance between being:
  - Easy to publish (the low hanging fruits), and
  - In high demand (so called “high-quality datasets”)

Having this balance increase the chance that the data will be reused immediately after its publication. The kind of quick success stories 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 for protection of personal data. Some government datasets (like statistical data) are based on data about the population. It is mandatory to use anonymization techniques to ensure that no personal data remains in the dataset before publication.

- Make sure that your Agency’s dataset does not contain personal data.
- If the dataset has been anonymized, make sure it cannot be de-anonymized.
- If you are not sure, it is recommended seeking expert advice (usually the National Statistical Office can provide).

There are established techniques for the anonymization of data. However, there are also methods to de-anonymize data. It is important to ensure that the method used for anonymization is an effective safeguard against de-anonymization.
5. Add context

Context will give re-users a better understanding of the OGD. 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, last modification and update frequency requirement

This information will help users understand the content of data; how to interpret it and if it meets users’ specific needs. More detailed documentation, user manuals, or methodologies can be linked from the metadata.
6. Structure data

To be used and reused, your Open Data should be available in well structured and in standard formats.

TO MAKE Your data processable by machine, select a solution that allows you to export your data in one or several of the following structured formats:

- XSL
- CSV
- JSON and GeoJSON
- RDF

*Unstructured information such as handwritten documents, or formats like Microsoft Word, PDF, or scanned images are 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. Use well established licenses for Open Data to maximise interoperability:

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

Don’t hesitate to use the latest version of Creative Commons licenses. They are suitable for content and for data, they are international and have localizations for many jurisdictions. It is okay to ask for attribution, but keep it as minimal as possible: e.g. name of your organization and a link to the license text.
8. Publish data online

Open Data are truly open when they are easily retrievable and accessible.

Initially it is fine making data available for download on a simple website or on a FTP server. However, this does not automatically mean that people can easily find this data.

- Make it available online
- Set up a portal
- Allow access to it through a non-mandatory sign-up

To increase accessibility, it is recommended publishing the data into a dedicated Open Data portal.

- Gather data in one single place
- Easily generate visualizations and APIs
- End-users will be able to readily browse through datasets
9. Ensure data accessibility and availability

- Consistently use URI’s for any published dataset, ensuring data can be accessed under the same URL on the Internet over time.

- Make datasets available for bulk download (for example as an export of an entire database).

- A great way to strengthen data availability is to generate Application Programming Interfaces (APIs).

APIs allow re-users to build applications around agency’s data. Thanks to APIs, every time you update the dataset, it will sync up with applications from third parties. APIs are specially 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 your data is online, the public needs to be informed about it.

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

Recommended modalities include:

- Using of social media to engage with potential re-users; and/or
- Organizing follow up roundtables on supply and demand to gather relevant feedback for future improvements.
11. Foster data reuse

Remember an Agency’s data can be reused to create social, economic and environmental value. Engage with potential re-users and intermediaries to help them making data useful for others. Invest time and resources in dedicated programmes that foster reuse uptake.

- Promote the reuse of your data via hackathons, workshops or conferences;
- Engage with new intermediaries to expand data outreach and enhance its usefulness;
- Foster reuse with dedicated programmes like mini-grants and fellowships, or business incubators, accelerators.

Go beyond the hackathon and partner up with others to set up dedicated programmes. These in turn support entrepreneurs and startups to engage with your Agency’s data to develop new services and applications. Every successful business that has been built on data is a success for the Agency as well.
12. Improve data skills

Data is not only useful 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 reuse it. However, government officials need adequate technical and data analytical skills to effectively use OGD.

- Introduce capacity building programmes to leverage technical and data analytical skills.
- Help other government agencies understand and use your Agency’s data.

Explore potential collaborations around capacity development to enhance data literacy, -management and -analysis within different types of organizations both inside and outside government.
13. Keep data fresh & preserved over time

Outdated datasets often become useless. It is critical updating data whenever possible and publishing it under a dedicated new version.

This does not mean duplicating older versions. In some cases it is essential that historical data remain available for comparison. It is highly recommended ensuring long-term preservation and availability over time.

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

Open Data is a long-term contribution to the community it serves. This requires long term commitment -- it takes time and energy but it won’t be long before your Agency can reap the benefits of its OGD.
Repeat the Sequence

Bravo! You’ve just successfully gone through the essential steps of identifying, prioritizing and publishing Agency data as Open Data.

Eventually, Open Data will get reused and might even get showcased on some famous data-related media.
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.