

# Checklist for opening data

This is a pdf summary of the operating model for data sharing. Find out more about the Operating model for data sharing from the avoindata.fi service at: [www.avoindata.fi/en/operating-model](http://www.avoindata.fi/en/operating-model)

## 1. Identify any restrictions or obligations

Public administration organisations have shared their data as open data, especially to fulfil obligations under Union legislation (including the INSPIRE Directive) or as a response to a request for information. In addition to EU obligations, Finland does not have more detailed national regulation on the opening of data.

When opening data, observe any legislation applicable to your field. If your organisation produces spatial data, for instance, the INSPIRE Directive may oblige you to publish it openly.

Involve the legal experts and data protection officer of your organisation in the process of opening data.

## 2. Plan the objectives, resource allocations and organisation of opening data

### **Organise the process of opening data:**

- Assign responsibilities.
  - Person(s) to whom responsibility for opening data has been clearly assigned
  - Information security and data protection experts
  - Technical experts
  - Communication expert
  - Analyst for data processing
- Arrange the necessary training.
- Plan the process.

### **In resource allocations, you should address the following:**

- Human resources
- Costs of processing the dataset
- Costs of dataset storage
- Costs of publishing the dataset and communication costs
- Costs incurred from maintaining and updating the dataset

It is advisable to start the opening of data with simple datasets, after which it is easier to proceed to more complex ones.

Ensure continuity in data opening and adequate resourcing by including data sharing in the organisation's strategy, performance targets and performance agreements.

Develop the process of opening data openly in cooperation with stakeholders.

### 3. Map suitable data for opening and user needs

To map the data to be opened, you can use some of the following:

- Information management model
  - Shows how information management has been organised in the information management entity (such as a municipality or government agency).
- Information management map
  - Describes public administration's shared information resources (name, purpose, data content and administrator) and their disclosure (what data is disclosed, to whom and for what purpose).
- Description to implement document publicity
  - Describes the data and case registers administered by an information management entity, such as a municipality.
- Data balance sheet
  - Describes such aspects as the status of information processing in the organisation and life cycle management in the processing of personal data.
- Interoperability method
  - Provides data definitions and metadata for digital services and information flows.

The opening of data should always be based on demand, because if data is not used, it also fails to create value. When you know what type of data your organisation has, determine the data for which there would be demand outside your organisation and what benefits could be gained from opening the data.

**When mapping user needs, the following may be useful:**

- Information requests and feedback received by the organisation
  - If certain datasets are requested frequently, it is advisable to publish them and give everyone open access to them if possible.
- User statistics of the organisation's website
  - If a topic attracts interest on the organisation's website, there may be demand for open data related to it.
- User surveys
  - Potential data users can be asked about the type of data they would find the most useful.

In addition to demand, also take the potential societal value of the data into consideration. If possible, prioritise the opening of datasets that promote equality, knowledge-based decision-making or the circular economy.

## 4. Prepare the data to be opened and select the form in which it will be shared

When you have identified the data suitable for sharing, assess the potential benefits, risks and costs associated with opening it. You can use the assessment tool found in the operating model.

### Prepare the data for opening:

- Find out who administrates and is responsible for the data and the information system in its background.
  - You can use such documents as the organisation's information management model for support.
- Consider questions related to copyrights, disclosures, data protection and information security.
  - For example: who holds the copyrights to the dataset, and can this data be published openly in general?
- Determine the completeness and level of accuracy of the data to be opened. This makes it possible to open the data without undermining either its potential usefulness or usability or, on the other hand, its information security.
- Find out if national or international standards exist for opening the data in question (for example for data modelling and formats) or if some other party has already opened similar data, in which case you could use its data model for opening the data.

### Ensure data protection:

- Determine if the dataset you are planning to open contains public data.
- Could the dataset contain personal data or information critical to the functioning of society?
- Together with the data protection officer of the organisation, assess the need for data aggregation and anonymisation.

### Decide if the data should be shared through an API or as a file or both:

- A file is a suitable way of sharing relatively small and static datasets that do not change much or often.
- Sharing data through an API is a good choice especially when the volume of the data is very large and the data is updated frequently or in real time.

### When selecting the form in which the data will be shared, note:

- The format in which the data can be exported from the information system: Can the data be shared through an API or as a batch file, or both?
- The type of resources and technical expertise that the organisation has: A precondition for sharing data through an API is that the organisation has technical expertise and the resources required to update the interface when necessary.

### Select a licence for the data that informs the user of the terms on which the data may be used:

- Use Creative Commons CC BY 4.0 or CC0 1.0 licences for open datasets.
  - Creative Commons Attribution International 4.0 (CC-BY) licence obliges the data user to acknowledge the original data source.
  - Creative Commons CC0 1.0 license imposes no restrictions on data use.

Make the decision to open the data, with the organisation's management if necessary, and manage residual risks, or risks that cannot be eliminated and that the organisation cannot, or chooses not to, counteract with measures.

## 5. Describe the dataset comprehensively and publish it in a data portal

By describing the data well, you will help data users understand your data and make its re-use easier.

Many data portals require some basic information about the data to be published, including the name of the dataset, the licence and the data administrator.

In addition to this basic information, the **metadata should include at least the following information:**

- description of the data content,
- data generation process,
- your assessment of the data quality.

If possible, also describe the data in Swedish and English, as this facilitates its international use.

**Publish the open data in a public data portal:**

- If the data to be published concerns the whole country or regions outside Helsinki Metropolitan Area, publish it in the [opendata.fi](https://opendata.fi) service.
- If the data to be published concerns an area in the Helsinki Metropolitan Area, publish it in the [hri.fi](https://hri.fi) service.

Inform potential users of the publication, for example on social media and in the organisation's newsletter.

## 6. When the data has been published, ensure that it is maintained

Strive to continuously improve the data quality. Encourage users to report any errors they detect, and develop the dataset based on feedback.

Update the dataset according to the update rate specified in the metadata, either automatically or manually.

Remember to update the metadata together with the dataset. Also keep older versions of the dataset available if they contain valuable data.

If the data published through an API is continuously updating, update the metadata when a new version of the interface is created, or when the data otherwise changes significantly.

Inform users about changes in the open data and interface.

Plan in advance what you will do if an error is detected in the published data.

**Depending on how serious the error is:**

- You can remove the dataset for the time being.
- If the error is not critical, you can report it in the metadata and announce the date on which the error will be corrected.
- Inform data users of the error.

## 7. Monitor the use of published data

The data published in a **dynamic format, usually through an API**, is relatively easy to monitor as analytics are often available on it.

On the other hand, monitoring the use of **static data (files)** is more challenging. While statistics can be obtained on downloads, for example, it is difficult to estimate the actual scope of use, as a file once downloaded can be used many times.

**You can monitor data use and its impact by:**

- collecting statistics on data downloads and, if possible, use cases.
- monitoring data use in new services and applications as well as on the media.
- conducting a cost-benefit analysis on the economic benefits of opening the data. The monetary value of many benefits is difficult to measure, however.
- commissioning reports or studies.

When assessing the impacts, it is advisable to take into consideration the entire life cycle of the data and the impacts of opening it from a wider societal perspective.

## 8. Support data users and actively collect feedback

Opening data alone is not enough, it is important to support the use of open data.

Interaction between data users and your organisation will increase the use of published data and is likely to also improve data quality.

- You should offer different feedback channels, including e-mail and a social media channel, through which data users can easily send feedback and development proposals.
- Work together with developers and organise activities associated with the data, such as developer meetings and hackathons.

## 9. Do not remove the dataset even if it is no longer updated

Datasets published as open data should be publicly available for as long as possible.

**When you stop updating the data:**

- Do not remove the dataset unless there is a good reason to do so.
- State in the metadata why and when updates were discontinued.
- You should note that even if the dataset is removed, data users can continue to use the data under the licence terms.