



User Guidelines

1. The purpose of Open Data Pakistan (ODP) is to create a culture of sharing data in an open format for the public good.
2. Only registered and approved organizations are able to access the platform as organization administrators to share data.
3. Please contact Open Data Pakistan to register as an organization through the connect page.
4. Organization members can:
 - View the organization's private datasets.
5. Organization editors can do everything as member plus:
 - Add new datasets to the organization
 - Edit or delete any of the organization's datasets
 - Make datasets public or private.
6. Organization administrators can do everything as editor plus:
 - Add users to the organization, and choose whether to make the new user a member, editor or admin
 - Change the role of any user in the organization, including other admin users
 - Remove members, editors or other admins from the organization
 - Edit the organization itself (for example: change the organization's title, description or image)
 - Delete the organization
7. Guidelines for sharing sensitive or private information and protected datasets
 - a) As part of the publishing process, data can be classified as per the following¹:
 - **Level 1 -Public.** Data available for public access or use.
 - **Level 2 - Internal Use.** Routine operating information for internal use; it is not proactively shared with the public. Use of level 2 data is intended for employees or a

¹ This is inspired from San Francisco's data classification standards https://sfcoit.org/sites/default/files/2019-09/DataClassificationStandard_FINAL_DRAFT.pdf

closed group in private mode. Certain data may be made available to external parties upon their request.

- **Level 3 -Sensitive.** Data regulated by legal regulations and privacy laws, or agreements such as contracts with non-disclosure agreements or other terms and conditions.
- **Level 4 -Protected.** Data that requires notifications to affected parties in case of a security breach.
- **Level 5 - Restricted.** Data with high impact and threat to human life or risk of an epidemic or catastrophic loss of major assets. This data must be verified by leadership or the concerned authorities for its classification.

- b) PII – “**personally identifiable information** refers to information that can be used to distinguish or trace an individual’s identity, either alone or when combined with other personal or identifying information that is linked or linkable to a specific individual. The definition of PII is not anchored to any single category of information or technology. Rather, it requires a case-by-case assessment of the specific risk that an individual can be identified. In performing this assessment, it is important for an organization administrator to recognize that non-PII can become PII whenever additional information is made publicly available (in any medium and from any source) that, when combined with other available information, could be used to identify an individual.”²
- c) PII data cannot be shared on the ODP.
- d) Data must be anonymized and aggregated to prevent release of sensitive, protected, and highly restricted data.
- e) Organizations may strategically publish private or sensitive data by performing the following methods³.

Method	What it is	Best for
Column Removal	Remove the privacy implicating columns. The simplest way to avoid any privacy issues, is to simply not publish the columns, which include private data.	Datasets with private or personal information that is not necessary for consuming and understanding the data.
	For example, if a dataset is a list of users and includes their name, address or other information, you can simply remove those columns from the dataset.	
Obfuscation	Mask or transcribe the data. Obfuscation can happen in a number of ways, but a common case is with address data. Sometimes we want to retain a proxy of the address without aggregating the data.	Datasets with private or personal information that is not necessary for consuming and understanding the data.

² Open Data Policy M-13-13 <https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>

³ <https://datasf.org/publishing/guidelines/>

Banding	Group the data. Banding is a way to obscure individual values. For example, instead of publishing age, you can publish age group. Other examples of banding include time (date to month to quarter).	Datasets where individual record data is important to publish but where too much detail can make it easy to identify individuals with uncommon mixes of characteristics.
Aggregation	Summarize the data based on a data property. Sometimes de-identifying the data is not sufficient. Your data might need to be aggregated either by geography or some other factor such as a category in the dataset.	Datasets where the individual records pose a privacy risk even if the identifying columns are removed. A common example of this is health related data. If the individual records (rows) are important to publish, use one of the other methods.

8. All data shared on this portal must be legally and ethically acquired, and properly sourced. Open licensing should be applied where possible.
9. Duplication of data should be avoided by checking open data already available on the portal.
10. Organizations can share data within their organization's network of members privately or can publish data publicly. Private data will never be shared with the public.
11. Users can share all types of data, raw, aggregated, structured, curated, uncurated, however, data in an open and machine readable format is preferable.
12. Data in open format is a file format with no restrictions, monetary or otherwise, placed upon its use and can be fully processed with at least one free open-source software tool.
13. Machine readable data is one that can be automatically read and processed by a computer, such as CSV, JSON, XML, etc. Machine-readable data must be structured data.⁴
14. Open Data Pakistan supports the following open file formats:

Type	Media	Description
CSV	Text	Comma-separated values
JSON	Text	JavaScript Object Notation
PDF	Binary	Portable Document Format
RDF	Text	Resource Description Framework
RSS	Text	RDF Site Summary/Really Simple Syndication
XLS	Binary	Microsoft Excel
XLSX	Binary	Microsoft Excel Open XML
XML	Text	Extensible Markup Language
ZIP	Binary	Typically contains a shapefile set (SHP, SHX, DBF)

15. Data will be held for an indefinite period unless the data administrator deletes it or there is an exceptional request to delete the data.

⁴ <http://opendatahandbook.org/glossary/en/>

16. Key to additional information or metadata for adding a dataset⁵:

Term	Definition	Comments
Title	A name given to the resource.	
Description	An account of the resource.	Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.
Category	Topic, sector or theme of the resource.	Agriculture, Food & Forests Cities & Regions Connectivity Culture Demography Economy & Finance Education Environment & Energy Government & Public Sector Health Housing & Public Services Manufacturing Public Safety Science & Technology
Tags	The topics of the resource.	Typically, the subject will be represented using keywords, key phrases, or classification codes. Recommended best practice is to use a controlled vocabulary.
Temporal coverage	Time period of resource. A point or period of time associated with an event in the lifecycle of the resource.	Date may be used to express temporal information at any level of granularity. Example, month year to month year pertaining to the data variables in the dataset.
Spatial coverage	Location of resource. The spatial applicability of the resource, or the jurisdiction under which the resource is relevant.	Spatial coverage may be used to express spatial information at any level of granularity. Spatial topic and spatial applicability may be a named place or a location specified by its geographic coordinates.
Organization name	Name of organization publishing the resource.	
Organization type		Federal government Provincial government Local government

⁵ <http://www.dublincore.org/documents/dces/?1401215562628>

Dataset type	Education Private NGO Other Geospatial Non-geospatial
Date created	Date resource was created.
Last updated	Date resource was last modified.
Source	Original source or link where resource was originally published or produced
Visibility	
Author	An entity (person, department or organization) primarily responsible for creating or producing the dataset. This could be the same as the publishing organization.
Publisher	An entity (person, department or organization) responsible for making the resource available. This could be the same or different from the author of the dataset.
Maintainer	Second point of contact responsible for the data.

17. Connect with us through the connect page if:

- You want to share a data story
- You want to register your organization
- You become aware of sensitive or private data that should not be shared publicly
- You become aware of duplication of data
- Data has been shared by a third party source, and the original source disagrees or has reservations, with a request to delete it.

18. The employees associated with ODP do not endorse or agree with the opinions expressed in the data shared on the portal

19. ODP can modify these terms or apply additional terms to reflect the changes.

20. These user guidelines are licensed under [Creative Commons Attribution-ShareAlike 4.0 International License](#).