# A Guide to Open Data

Open Nepal

As global trends towards transparency and accountability gain ever-increasing traction, the phrase "**Open Data**" is becoming increasingly used - both internationally and within Nepal. But, what is **Open Data** and what exactly does it mean for Nepal? This document answers some basic questions, and dispels some myths, about the increasing demands and challenges of ensuring 'open data'.

For more information about **open data** please contact Open Nepal, at info@opennepal.net, or visit www.opennepal.net

## What is Open Data?

The term **open data** simply refers to raw data that is made available to the public in a way that can be accessed and processed by computers. By disseminating data in this way, data can be freely used, analysed and shared by anybody who might want to use it for any purpose.

In essence, open data is a tool for delivering greater equality of access to information and greater opportunity for innovation. By enabling improved availability and accessibility of information, open data can empower government, citizens, business and civil society to engage with public issues and make better informed decisions to improve their lives.

## What is open government data?

The term "open government data" is used to describe data produced or commissioned by government that has been publicly shared in an open format. The publishing of government data presents opportunities for the government as a whole to be more efficient, effective and accountable, and allows for better-informed decision making. Open government data includes a wide range of datasets, from census data to budget data, and data on public services to environmental data, etc.

Some governments have made their data open by uploading their datasets onto online portals accessible to the public. A large number of countries now have open data portals - over 500 portals from countries, districts and states can be seen on http://dataportals.org/.

For example, the UK government's portal (https://www.data.gov.uk/) releases data to help citizens understand how government works and how policies are made. The Kenyan government's open data portal (https://opendata.go.ke/) provides citizens with full access and rights to its data, including government, social, demographics, geospatial, environment, infrastructural, agricultural, transportation and economic data.

## Why is open government data a good thing?

The benefits gained from the sharing and use of open data by government, the private sector and citizens are numerous and well-established. For example:

- **Improved transparency:** Open data allows all parts of a government to know what other departments and agencies are doing. Equally, it also allows public-spirited citizens to know what their government is doing, how it is spending their money and provides them with the material needed for them to engage effectively in government decision-making and the formation of public policy. Given the rapid expansion of information technology, the public access, analysis and sharing of government data provides a solid foundation for a functioning, transparent and democratic society.
- **Improved efficiency:** Open data can enable better coordination within government agencies by making data easier to access, analyse and combine across different sections of government. Efficiency gains can be experienced through a resulting improvement in the delivery of public services, greater awareness of government expenditures and more collegial and inclusive decision making.
- **Improved innovation:** Open data is a key resource for stimulating social and commercial activities. Using open data, such as agricultural, pricing and transportation data, technical experts can build applications that provide useful information for government and citizens. By opening up data, government can therefore stimulate entrepreneurship and drive business innovation, which will create new technologies, domestic job expansion and greater commercial and social value.

## Dispelling some myths about open data

There are a number of common misconceptions about open data, which are addressed here:

1. **Open data does NOT mean that original data-points can be altered by the user** – open data does not allow individual data points to be changed on the original source of publication, e.g. the host website or data portal.
2. **Open data does NOT mean that personal data becomes public** – the focus of government open data portals is non-personal data. Open data does not contain information about specific individuals, and identifying features can be removed before data is released publicly.
3. **Open data is NOT a breach of national security** – the data supplier, often a government, can chose which data they wish to make open. National security restrictions may apply to some types of data.
4. **Open data is NOT necessarily expensive or difficult** – Many off-the-shelf frameworks exist already to support data sharing, and modern cloud-based platforms can prevent the need for large-scale hardware investment.
5. **Data presented in PDF reports or written as website text is NOT open** – for data to be considered open it has to be shared in a machine readable format.

| Is it open? | |
|---|---|
| PDF | No |
| Website text | No |
| Excel file | Yes |
| Comma separated files | Yes |

## When can data be considered "open"?

For data to be considered open, it must be in a machine readable format (e.g. Jason, XML, CSV), ideally free to download, and provided under terms that permit reuse, redistribution and intermixing with other datasets. A number of globally recognised principles have been developed that determine whether data is open.

For example, the Sunlight Foundation states that government data shall be considered open if it that complies with the principles below[i]:

1. **Completeness:** Datasets released are the entirety of what is recorded about a particular subject, including metadata, with the exception of personally identifiable information.

2. **Primacy:** Datasets released are the primary data collected at the source, with the finest possible level of granularity, not in aggregate or modified forms.

3. **Timeliness:** Datasets released are made available as quickly as necessary to preserve the value of the data.

4. **Ease of Physical and Electronic Access:** Datasets released are as accessible as possible, though physical or electronic means.

5. **Machine Readability:** Datasets are reasonably structured to allow automated processing, and shared in widely-used file formats that can be easily machine processed.

6. **Non-Discrimination**: Datasets released are available to anyone, with no restrictions or requirement of registration.

7. **Use of Commonly Owned (Open) Standards:** Datasets are available in a format over which no entity has exclusive control.

8. **Licensing:** Datasets are not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed as governed by other statutes.

9. **Permanence:** Datasets are available online in permanent archives, with appropriate version tracking over time.

10. **Usage Costs:** Datasets are available at no or minimal charge to the user.

**What could open data mean for Nepal?**

Recent events in Nepal have demonstrated the value of open data for supporting decision-making and informing policy. The importance of open data became particularly apparent following the 2015 earthquake.

For example, open government data about the demographic, geographic and economic composition of wards would have allowed government and humanitarian agencies to make better-informed decisions about how much food to distribute and how best to reach earthquake-affected villages.

The Government of Nepal's Disaster Risk Reduction Portal provided some much needed information on people, livestock, health facilities, schools and buildings, however had this information been released as open data it would have been more easily used.

For example, access to the 2011 census provided useful information regarding population size and demographics, however much of the information was incomplete or released only in PDF format which made it difficult for agencies responding to the earthquake crisis to use effectively and efficiently…

## Useful further information:

**Is it open according to international standards?**

| Dataset | Yes/No | Why? |
|---|---|---|
| Red book data shared on the Ministry of Finance's website | No | • Only in PDF form<br>• Incomplete dataset<br>• Not released in timely manner |
| Data shared on Kathmandu Ward 7 citizen's portal (http://model4g.net/) | Yes | • Budget data<br>• Machine readable format |
| National Population & Housing Census 2011 - Central Bureau of Statistics | No | • Only in PDF form<br>• Not released in timely manner |
| Nepal Earthquake Portal (http://earthquake.opennepal.net/) | Yes | • Humanitarian flow data<br>• Machine readable format |

- Open Data Handbook: http://opendatahandbook.org/
- Open Government Data Handbook: https://opengovdata.io/
- Introduction to Open Government Data: http://opengovernmentdata.org/
- The 8 Principles of Open Government Data: http://opengovdata.org/

Open Nepal is a network of organisations working to improve development outcomes through the sharing and use of open data. Together we aim to build an inclusive information system of data suppliers and users, including government, civil society, media and the private sector.
If you would like to learn more about open data please visit the Open Nepal website: www.opennepal.net, follow us on twitter @Open__Nepal or contact us at info@opennepal.net.

[i] http://sunlightfoundation.com/policy/documents/ten-open-data-principles/