Understanding Government:
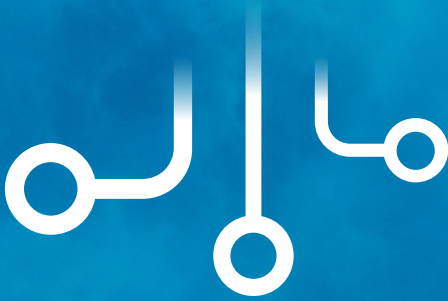# Improving Institutional Transparency and Accountability Through Open Data

Handbook

Table of Contents

What is Open Data?

We are living in the information age. IBM assert that every day, we create 2.5 quintillion bytes of data — so much that 90% of the data in the world today has been created in the last two years alone. Governments are no different to individuals and private organisations in that they also produce boundless amounts of data as part of their day-to-day work — sensors used to gather environmental data, public-transport passenger data, hospital attendance figures, school exam results etc. This data is used by governments to deliver and improve public-services, and to data-driven policy-making. However, in the public sector today, it is usually this secondary information that is published: the studies, reports, tools, policies, etc. Although all content can be made Open, these examples of public documents and tools would not be considered as Open Data. Open Data rather refers to raw data or facts, such as numerical or statistical data. An example of raw data is a temperature measurement, the year a school was built, the modes of transport available in a city, etc. Conclusively, the idea behind Open Data is to publish the raw data, as this facilitates the reuse of the data for new purposes, bringing both societal and economic benefit.

The Open Knowledge Foundation defines Open Data as: A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike. Open Data should be free to be used both non-commercially and commercially. The benefits of publishing Open Data are as follows:

When publishing data, the five main goals are:

> *That the data is useful, meaningful and accessible.*

> *That the data published brings value, provides ongoing dialogue and discussion to any economy.*

> *That the data is transparent and accountable, which helps rebuild trust between government and the public.*

> *That citizen participation through discussions around data and uses of the data increases, for example through apps.*

> *That activities, policies and m easures of public institutions are more transparent and subject to greater scrutiny by citizens, civil society and the media.*

> *That the data from researchers, developers and entre-preneurs, is being analyzed and reused in new and innovative ways.*

Open Data Internationally

Elements of Open Data can be traced back to many different initiatives and movements, such as:

> *Freedom of Information*
> *Transparency and participation initiatives*
> *Data exchange initiatives*
> *Reuse of Public Sector Information*
> *Open Access*
> *Open Source*
> *Open Government*

The concept of Open Data in the sense we use it today emerged in around 2007, with most of the initial activity happening in the U.S.A and in the UK.

In the past years, local, regional, and national governments have developed Open Data strategies, set up Open Data portals, and published their data under Open Licenses. Many non-governmental organisations also now publish Open Data, for example, the World Bank and the United Nations. Nowadays, some commercial entities are even starting to publish Open Data.

There are a number of organisation and groups that are driving Open Data research, best practice and technologies internationally. These include, but are not limited to the European Commission, the World Bank, the Open Government Partnership, the Open Data Institute, the Open Knowledge Foundation, W3C, the Insight Centre for Data Analytics, the Open Data Research Network, Govlab and the Sunlight Foundation.

Open Data in Kosovo

The Open Data Inventory (ODIN) assesses the coverage and openness of official statistics to help identify gaps, promote open data policies, improve access, and encourage dialogue between national statistical offices (NSOs) and data users. ODIN 2017 includes 180 countries, including most all OECD countries.

Kosovo is ranked 68th in the 2017 Open Data Inventory, which is an improvement in comparison to 2016's measurement that ranked the country as 82nd. The country's ODIN score improved from 39 to 44 as well. In South East Europe, Albania was ranked 40th and Macedonia was ranked 32nd. Others, such as Serbia and Bosnia were ranked 94th and 109th.

The nonprofit organization, Open Data Kosovo, champions the opening of governmental data in Kosovo since its founding in 2014. Most of the data liberated by the government of Kosovo as open data has thus far took place under the leadership and expertise of ODK, including procurement, air quality, water surface quality, election monitoring, asset declaration and other types of data. ODK advocates and lobbies for government to adopt digitally-driven transparency best practices, such as hosting and populating a national data portal (Government Open Data Portal) and using open licensing for data and software source code.

Government Open Data Portal was developed to serve as a tool for government officials to enable transparency through the publication of government datasets. Institutionalizing the usage of the data portal was an important success factor for this project and a comprehensive training programme aimed at government officials was also designed and implemented.

Other good example of Open Data in Kosovo include Gjakova Portal. Through this portal the municipality of Gjakova is closer to its citizens and presents the past and the current situation of the city, providing the necessary information for every field of interest. The portal offers the opportunity to promote Gjakova in the fields of investment, tourism and entertainment, offering direct and easy access to any information you need about Gjakova, within and outside the territory of Kosovo.

Open Data Standards

Data standards help give a common meaning to data. This is especially important when data is shared across departments, organisations, or even international boundaries, so that the data user can interpret the true meaning of the data. Data standards not only define the meaning of certain concepts, but also how concepts relate to each other. Data standards should have both a human-readable and machine-readable representation.

Data standards allow smooth data exchange within a certain community — the community of adopters and implementer. These communities can vary in size. For some standards, such as the aforementioned ISO standard, it is a global community. And most data standards address a community of specialists that are interested in exchanging a particular kind of data. For example, there are standards for geospatial data (developed by users and vendors of geospatial information system (GIS) software), for statistical data (developed by statistics offices and other statistics producers), for financial data, etc.

Open Data Licensing

What makes Open Data 'open' is that it is free to be used, including for commercial use. Associating an Open License with Open Data is necessary to ensure the legal grounding for its potential reuse. For a data user (individual/ organisation/company/etc.) wishing to use and build on top of public data, they require assurance of what they legally can and can't do with the data. If no license is specified, each data-user must contact the data publisher on a case-by-case basis. For a data publisher an Open License allows them to define the terms under which the data can be used. The main terms that are specified in Open Licenses are attribution and share-alike. Generally speaking, attribution means that the data user must credit the data publisher for the original creation and share-alike means that the data user must license their new creations under identical terms as the original data. Attribution can help boost visibility of Open Data initiatives.

In addition to publishing the data under an Open License, the metadata should also be made available under an Open License. This ensures that data is discoverable and increases its potential use as people can freely read descriptions of data.

Metadata should always be published under an Open License, even if the underlying data is not (Metadata means structured information about a resource, for example, meta-data about a dataset includes its title, description, theme, creation data, temporal coverage, etc)

All commonly used licenses for Open Data declare that data is provided as-is. The publisher does not guarantee the data's correctness, timeliness, its provision in a particular format, or its future availability. Anything is subject to future change or correction.

While it is best practice to publish Open Data free-of-charge, there can be justification to charge a fee for SLAs if they require the data publisher to provide additional data services and/or infrastructure on top of the original data.

5 ★ Open Data

In 2010, the inventor of the World Wide Web and the creator and advocate of the Semantic Web and Linked Data, Sir Tim Berners-Lee, suggested a 5-star deployment scheme for Linked Open Data. The idea behind this format is that the greater the number of stars, the more reusable the data is, and the easier it is to reuse and interconnect data. The rating begins at one star and data gets stars when proprietary formats are removed and links are added. Below, we show what each star represent:

One-star web data ★

As a CONSUMER

- You can look at it
- You can print it
- You can store it locally (on your hard drive or on an USB stick)
- You can enter the data into any other system.
- You can change the data as you wish
- You can share the data with anyone you like

As a PUBLISHER

- It's simple to publish
- You do not have explain repeatedly to others that they can use your data

Two-star web data ★★

As a CONSUMER

You can do all what you can do with one-star Web data and additionally:

- You can directly process it with proprietary software to aggregate it, perform calculations, visualise it, etc
- You can export it into another (structured) format

As a PUBLISHER

- It's still simple to publish

Three-star web data ★★★

As a CONSUMER

You can do all what you can do with one-star Web data and additionally:

- You can manipulate the data in any way you like, without the need to own any proprietary software package

As a PUBLISHER

- You might need converters or plug-ins to export the data from the proprietary format
- It's still rather simple to publish

Four-star web data ★ ★ ★ ★

As a CONSUMER

You can do all what you can do with three-star Web data and additionally:

- You can link to it from any other place (on the Web or locally)
- You can bookmark it
- You can reuse parts of the data
- You may be able to reuse existing tools and libraries, even if they only understand parts of the pattern the publisher used
- Understanding the structure of an RDF "Graph" of data can be more effort than tabular (Excel/CSV) or tree (XML/-JSON) data
- You can combine the data safely with other data. URIs are a global scheme so if two things have the same URI then it's intentional, and if so that's well on it's way to being 5-star data!

As a PUBLISHER

- You have fine-granular control over the data items and can optimise their access (load balancing, caching, etc.)
- Other data publishers can now link into your data, promoting it to 5 star!
- You typically invest some time slicing and dicing your data
- You'll need to assign URIs to data items and think about how to represent the data
- You need to either find existing patterns to reuse or create your own

Five-star web data ★ ★ ★ ★ ★

As a CONSUMER

You can do all what you can do with four-star Web data and additionally:

- You can discover more (related) data while consuming the data
- You can directly learn about the data schema
- You now have to deal with broken data links, just like 404 errors in web pages
- Presenting data from an arbitrary link as fact is as risky as letting people include content from any website in your pages. Caution, trust and common sense are all still necessary

As a PUBLISHER

- You make your data discoverable
- You increase the value of your data
- Your own organisation will gain the same benefits from the links as the consumers
- You'll need to invest resources to link your data to other data on the Web
- You may need to repair broken or incorrect links
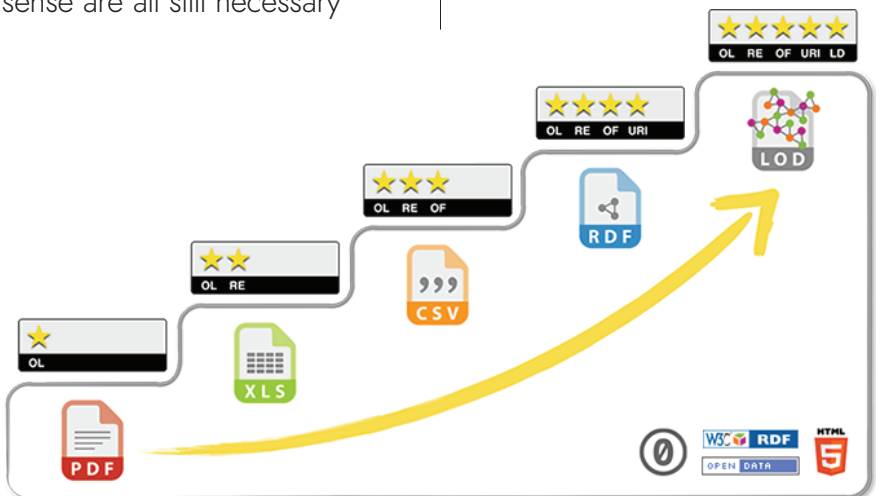
Open Data Principles

The act of providing open data is to publish and share, and it can be understood that this involves responsibility and due diligence. When publishing data openly, the aim is to ensure it is of a high enough quality to be accessed, used and understood. It is the intention of this policy to reduce potential issues that make data less useful.

Principle 1

Principle: Open data should respect individual's privacy.
Best practice: Ensure that open data is free from information that can be used on its own or with other information to identify a living person, as outlined in the ICO code of practice.

Principle 2

Principle: Open data should be comprehensive for the subject.
Best practice: Be sure to quality assure the data, in terms of the level of completeness and readiness for publication. Do not knowingly publish data that is incomplete for the relevant focus and/or time period.

Principle 3

Principle: Open data should be relevant and succinct for the subject.
Best practice: Always consider the scope and spread of the data - to make it useful for those who may want to access it.

Principle 4

Principle: Open data files should be a workable size that can be used on a personal computer.
Best practice: Always consider the size of the data - to make it usable for those who may want to access it.

Principle 5

Principle: Open data should use commonly available references where applicable.
Best practice: Always provide codes and lookups for our data, particularly in terms of administrative geography. We will publish data that is free from jargon and document acronyms and abbreviations used.

Principle 6

Principle: Open data should be presented in a common technical file format.
Best practice: Publish data in common, accessible and standard formats such as plain text for example, CSV, JSON and XML. Do not publish open data in bespoke, redundant or proprietary formats

Principle 7

Principle: Open data should be appropriately licensed.
Best practice: Always issue an open licence with open datasets.

Principle 8

Principle: Open data should be well documented.
Best practice: Always provide notes and guidance with datasets. Always detail the nature, scope and purpose of the open datasets in a release document, available to all. Do not purposefully provide data that is poorly described, or requires sector knowledge to comprehend.

Principle 9

Principle: Open data should be accessible and consistent.
Best practice: Always publish data consistently, making it accessible to all. Do not publish data that needs passwords or put other access restrictions in place, unless these are signposted.

Principle 10

Principle: Open data should be timely.
Best practice: Always provide regular and timely updates to relevant open datasets. Do not miss updates to relevant datasets, or let data go "stale".

Principle 11

Principle: Welcome feedback and discussion about open data.
Best practice: Always make it clear how to provide feedback on open data, and where appropriate, publicly document actions. Do not publish data without a feedback mechanism.

Principle 12

Principle: Use own open data.
Best practice: Use own datasets in internal and external channels - providing commentary, analysis and insight.

Principle 13

Principle: Use other relevant open data.
Best practice: Act to utilise other openly available datasets in your work. Always provide clear attribution and guidance to the source of the data, and any actions you may have taken.

Principle 14

Principle: Take down inaccurate open data.
Best practice: Act to remove or replace published open data where mistakes have been identified in a timely and considerate manner.

Open Data Kosovo
Who we are ?

Open Data Kosovo designs, develops and delivers scalable data-driven solutions for a diverse portfolio of clients, among whom Kosovo's public institutions occupy a special place. ODK's expertise in data collection and analysis provides its clients and partners with the tools and insights that support data-driven decision-making, policy-making, and process optimizations in day to day strategies.

ODK has a strong commitment to gender equality, focusing its youth engagement activities towards empowering young women interested in technology and partnering with relevant grassroots organizations such as Girls Coding Kosova.

All local projects undertaken by ODK involve making government processes more transparent by opening government data in a way that sheds light on the internal workings of government activities (e.g. opening procurement data to better document and monitor procurement practices)

Outside of Kosovo, ODK leads and partners in several initiatives that promote open data best practices, among others the Open Contracting Data Standard.

It has received international acclaim for developing a red-flag algorithm that automatically analyses the open procurement data in Kosovo and highlights possible irregularities in uses of public money. At the cutting-edge of open data thinking in the Western Balkans and worldwide, ODK has contributed to the Aspen Institute conference "The Role of Digitalization for the Economic Development of the Western Balkan Region" with a research paper and presentation on "Citizen Engagement in Kosovo with Open Data Driven Civic-Tech Initiatives for Increased Government Transparency and Youth Employability."

As part of of TransparenCEE, a network of Central Eastern European and Eurasian NGOs interested in using technology in transparency and accountability work, ODK has conducted extensive research into openness of all public institutions in Kosovo (including thousands of indicators) and drafted policy reports spanning the entire institutional spectrum. It has contributed to the production of a regional government openness index that allows for cross-country comparative analysis.
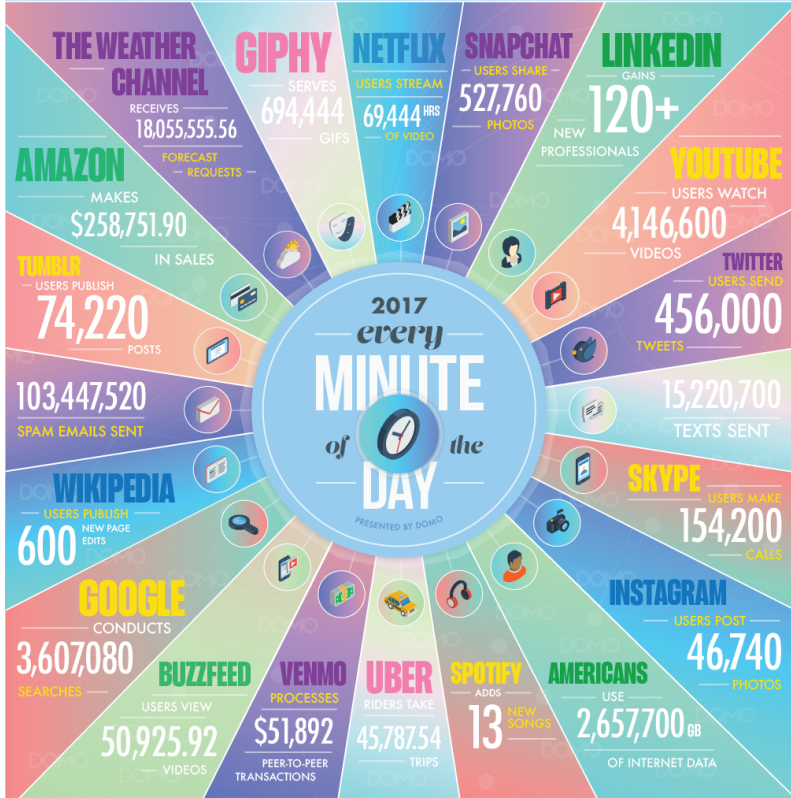
ODK is also involved in providing digital product development services to international NGOs that operate in areas concerned with social good and humanitarian interventions. In this field, ODK received the Albanian ICT Award 2017 recognition for Decode Darfur, a project (application) that enables identification of areas affected by violence and destruction in Sudan through the use of satellite imagery.
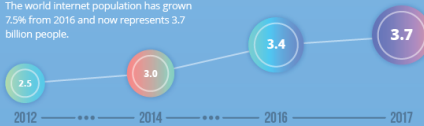
DOMO

DATA NEVER SLEEPS 5.0

How much data is generated *every minute?*

90% of all data today was created in the last two years—that's 2.5 quintillion bytes of data per day. In our 5th edition of Data Never Sleeps, we bring you the latest stats on just how much data is being created in the digital sphere—and the numbers are staggering.

THE WEATHER CHANNEL
RECEIVES
18,055,555.56
FORECAST REQUESTS

GIPHY
SERVES
694,444
GIFS

NETFLIX
USERS STREAM
69,444 HRS
OF VIDEO

SNAPCHAT
USERS SHARE
527,760
PHOTOS

LINKEDIN
GAINS
120+
NEW PROFESSIONALS

AMAZON
MAKES
$258,751.90
IN SALES

TUMBLR
USERS PUBLISH
74,220
POSTS

103,447,520
SPAM EMAILS SENT

WIKIPEDIA
USERS PUBLISH
600
NEW PAGE EDITS

GOOGLE
CONDUCTS
3,607,080
SEARCHES

BUZZFEED
USERS VIEW
50,925.92
VIDEOS

VENMO
PROCESSES
$51,892
PEER-TO-PEER TRANSACTIONS

UBER
RIDERS TAKE
45,787.54
TRIPS

SPOTIFY
ADDS
13 NEW SONGS

AMERICANS
USE
2,657,700 GB
OF INTERNET DATA

2017
every
MINUTE
of **the**
DAY
PRESENTED BY DOMO

YOUTUBE
USERS WATCH
4,146,600
VIDEOS

TWITTER
USERS SEND
456,000
TWEETS

15,220,700
TEXTS SENT

SKYPE
USERS MAKE
154,200
CALLS

INSTAGRAM
USERS POST
46,740
PHOTOS

The world internet population has grown 7.5% from 2016 and now represents 3.7 billion people.

2.5 — 2.0 — 3.4 — 3.7
2012 ••• 2014 ••• 2016 — 2017

GLOBAL INTERNET POPULATION GROWTH 2012–2017
(IN BILLIONS)

With each click, swipe, share, and like, businesses are using data to make decisions about the future. Domo gives everyone in your business real-time access to data from virtually any data source in a single platform for smarter decision-making at any moment.

Learn more at domo.com

DOMO

References

5 star Open Data, 2015
www.5stardata.info/en

Domo, Inc. 2018
www.domo.com/learn/data-never-sleeps-5

Gjakova Portal, 2018
www.gjakovaportal.com/en/About-Us

Ontotext, 2018
www.ontotext.com/knowledgehub/fundamentals/five-star-linked-open-data

Open Data Kosovo, Retrieved 2018
www.opendatakosovo.org

Open Data Inventory, 2018
www.odin.opendatawatch.com

Open Data Ireland, "Best Practice Handbook", 2014
www.per.gov.ie/wp-content/uploads/Best-Practice-Handbook.pdf

Open Government Data Portal, 2016
www.opendata.rks-gov.net

Open Data Services Co-operative Limited, 2015
www.opendataservices.coop