# Open Government Data Ecosystems

Handbook on Open Data

Handbook on Open Data

Every day we are surrounded by data, let that be how the alarm clock works when we wake up in the morning up to the weather weekly predictions we check by the end of the day. In the digital world we are living, the importance of data has grown exponentially. Simultaneously, the need to truly understand data and how data applies to different fields in life has risen altogether. But, what is open data? What is the history and journey of open data? How do we know the principles and standards of open data? What are the 5 Star Data? Many more questions are inevitable to truly understand the nature of data and finally our rights to access such data. Thus, this handbook will provide an overview on open d : The concept of Open Data, History of Open Data, Open Data Principles, Open Data Standards, 5 Star Data, Licensing of Open Data, Applications based on Open Data, Data Analysis, Data Science, Data Visualization, Open Contracting Data Standards, Open Data Global Initiatives, and Open Data Initiatives in Kosovo.
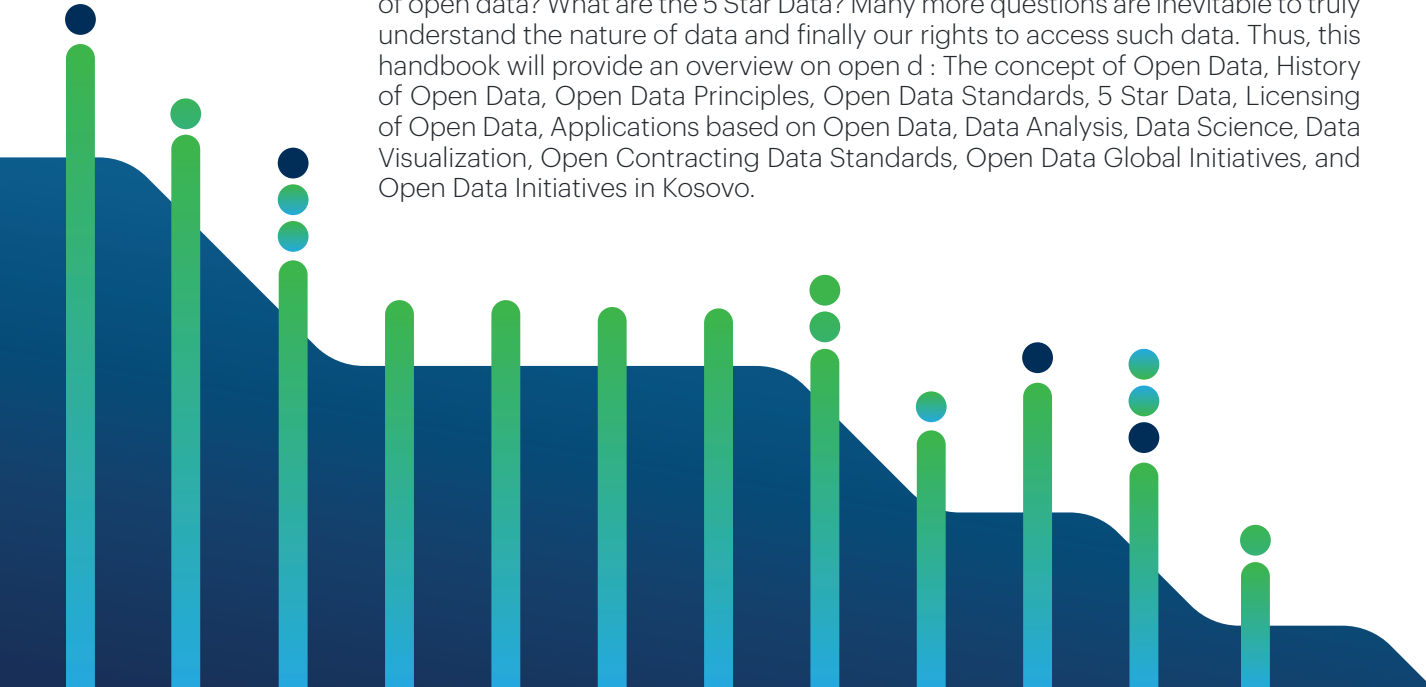
1.   What is Open Data?

To practically understand what open data is, we will use the definition provided by the Open Knowledge Foundation, which is: "Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share-alike."[1] In order to best define open data, there are three main goals which make it easier for us to grasp it:

- Availability and Access: the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.

- Re-use and Redistribution: the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.

- Universal Participation: everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavor or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

---

1   https://opendefinition.org/
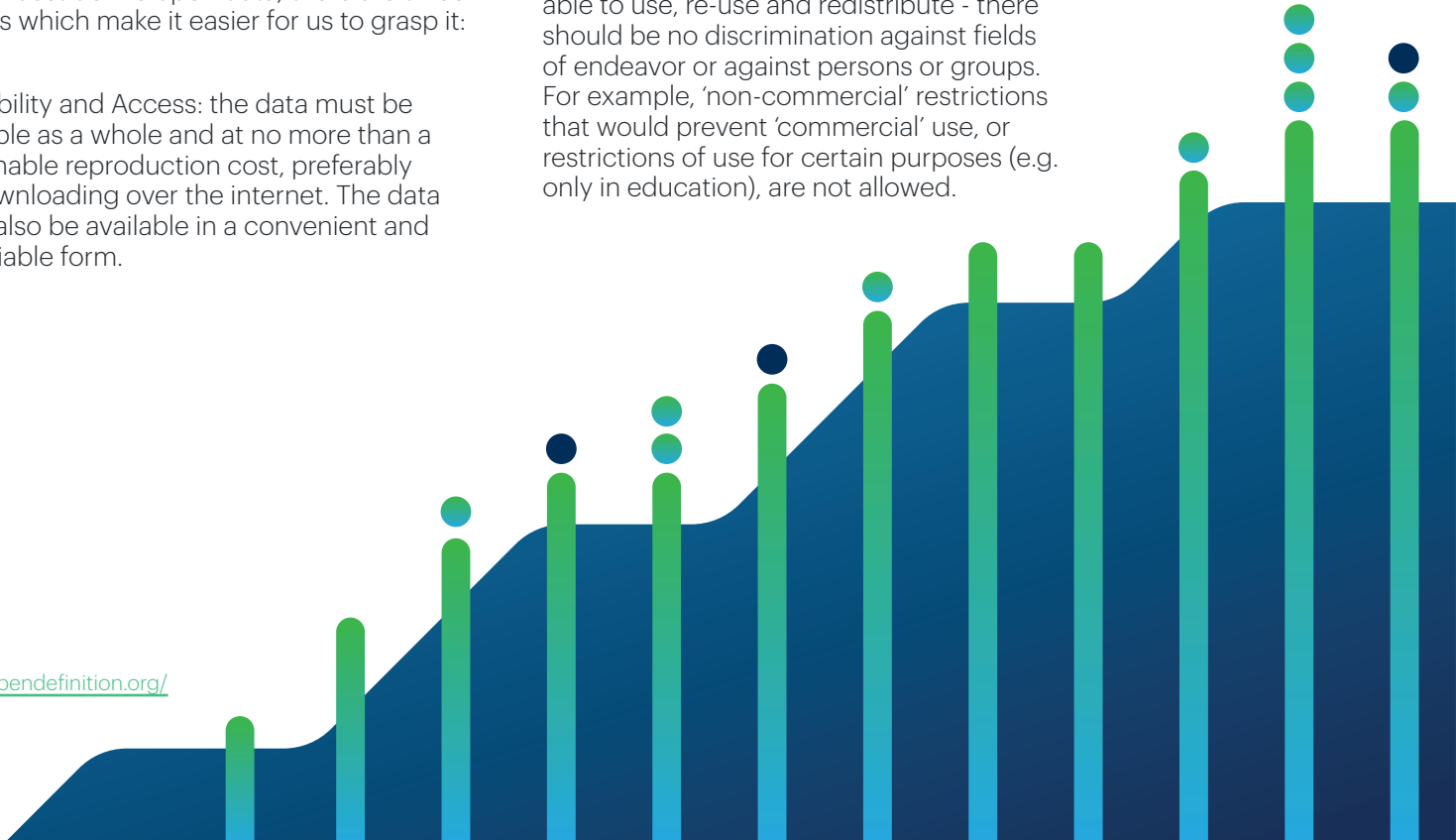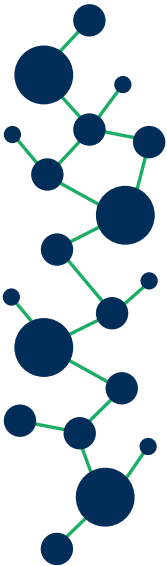
2. History of Open Data

In order to understand the concept of open data more deeply, the sole definition of it does not serve the purpose. Instead, knowing the history behind open data might help us understand better to then draw more accurate analysis while utilizing open data tools and mechanisms. As such, this part of the handbook will provide the history of open data, how the concept emerged, and the journey it encountered up to today.

According to Simon Chignard from the Paris Innovation Review, "The term open data appeared for the first time in 1995, in a document from an American scientific agency. It dealt about the disclosure of geophysical and environmental data."[2] Mr. Chignard then continues to explain the history of open data in time. In December 2007, thirty thinkers and activists of the Internet held a meeting in Sebastopol, north of San Francisco. Their aim was to define the concept of open public data and have it adopted by the US presidential candidates.

Among them were two well-known figures: Tim O'Reilly and Lawrence Lessig. The first is familiar to the techies: this American author and editor is the originator of many vanguard computer and Internet movements; he defined and popularized expressions such as the open source and Web 2.0. Lawrence Lessig, Professor of Law at Stanford University (California), is the founder of Creative Commons licenses, based on the idea of copyleft and free dissemination of knowledge. Then, Mr. Chignard continues by explaining how a little over a year later, President Barack Obama took office in the White House and signed three highly important presidential memorandums, two of which concerned open government which is mostly based on open data as a key pillar. As such, the founding principles such as: transparency, participation and collaboration came into use.

2 http://parisinnovationreview.com/articles-en/a-brief-history-of-open-data#:~:text=Open%20data%20responds%20to%20a,of%20geophysical%20and%20environmental%20data.

The Open Data initiative in Kosovo started back in 2014, with a close collaboration between Kosovo Foundation for Open Society, Open Data Kosovo and Ministry for European Integration, in cooperation with other stakeholders such as the United Nations Development Programme and other civil society organizations. The initiative took part under the flagship of ODK to develop the first Open Data Portal for the Kosovo government. The first result of this strong partnership was achieved with the development of the first Open Data Portal by Open Data Kosovo back in May 2016, an initiative of the Ministry of European Integration together with the Ministry of Public Administration and the Agency of Information Society in Kosovo. Then, by end of May 2016 the Kosovo Open Data Charter was approved by a verdict of the Government of Kosovo, a verdict which delegated the open data executive power to the Ministry of Public Administration to coordinate the process and further develop the initiative in close collaboration with other stakeholders including CSOs and private organizations by launching new projects into it. The Ministry of Public Administration with its umbrella Agency for Information Society with the methodology of the World Bank published the Open Data Readiness Assessment (ODRA). The ODRA report set forward the necessity of cross-collaboration between stakeholders and the creation of the open data steering committee was created, a committee in which Open Data Kosovo is part of. Moreover, it sets forward that each of the public institutions have a designated person which will be responsible for publishing the data. By July 2019 the Open Data Charter became part of the Law on Access to Public Documents with its own chapter where such an integration of the open data initiative aims to ensure a widespread request also from citizens for information to be published in an open format.

3.   Open Data Principles

There are 8 principles of Open Government Data which were initially authored by a group convened by Carl Malamud in California back in 2007. According to Joshua Tauberer writing for the Open Government Data book, even though these principles could be considered outdated, their relevance did not change through the years. To go on, the 8 open government data principles are as following:

• Data Must Be Complete; All public data are made available. Data is electronically stored information or recordings, including but not limited to documents, databases, transcripts, and audio/visual recordings. Public data are data that are not subject to valid privacy, security or privilege limitations, as governed by other statutes.

• Data Must Be Primary; Data is published as collected at the source, with the finest possible level of granularity, not in aggregate or modified forms.

• Data Must Be Timely; Data is made available as quickly as necessary to preserve the value of the data.

• Data Must Be Accessible; Data is available to the widest range of users for the widest range of purposes.

• Data Must Be Machine Processable; Data is reasonably structured to allow automated processing of it.

• Access Must Be Non-Discriminatory; Data is available to anyone, with no requirement of registration.

• Data Formats Must Be Non-Proprietary; Data is available in a format over which no entity has exclusive control.

• Data Must Be License-free Data 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.[3]

---
3   https://opengovdata.io/2014/8-principles/

4. Open Data Standards

Data standards help give a common meaning to data. This is especially important when data is shared across departments, organizations, 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 implementers. 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.

5.   5 Star Data

The concept of 5 Star Data is a 5-star deployment scheme that needs to be covered in order to understand the concept of open data taking a more in-depth approach. Initially, the inventor of the Web and the initiator of the Linked Data - Tim Berners-Lee suggested the development of such a 5-star deployment scheme for Open Data.[4]

Generally, the concepts of all 5-star data are as follows: 1-star data; make your stuff available on the Web (whatever format) under an open license, 2-star data; make it available as structured data (e.g., Excel instead of image scan of a table), 3-star data; make it available in a non-proprietary open format (e.g., CSV instead of Excel), 4-star data; use URIs to denote things, so that people can point at your stuff, and 5-star data; link your data to other data to provide context.

In order to best understand what such data explanations mean, it is helpful if we elaborate for each of the 5-star data taking into consideration the point of view, namely, what such a scheme means for users and for publishers.

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 aggre- gate it, perform calculations, visualise it, etc
- You can export it into another (structured) format

As a PUBLISHER
-It's still simple to publish

4   https://5stardata.info/en/

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 (relat- ed) 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 discover- able
- 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

6. Licensing Open Data

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/organization/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 the 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, metadata 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.

7.   Applications based on data

The Future Workplace
(https://thefutureworkplace.org/)
TFW tool which aims to analyse the ICT market information in order to improve the development of the sector, and to provide recommendations on the development of other sectors. Moreover, the data driven from the ICT sector will serve as a starting point, to be further used in job creation and employability of youth in other industries with high potential in Kosovo. The digital tool will obtain data from the education sector including private and public high schools, vocational schools, and universities in order to analyse past and current trends, and will serve as a start point on building the same solutions for other sectors. Moreover, the digital tool will analyse the gender differences in regards to education and percentage of employment in the job market.

OPEN BUSINESSES
(www.biznesetehapura.com)
Open Businesses is a business registration search engine with data from over 170,000 businesses in Kosovo. The data was scraped from the Kosovo Business Registration Agency (ARBK). Opening business data allows economic analysis of sectors as well as various economic activities. It creates space for analyzing economic trends in different periods and regions. It creates space for market analysis by local and international investors. Opening these business data enables data linking with other sectors such as public procurement or property declaration, which would strengthen anti-corruption mechanisms.

Open Contracts
(https://kontratat.rks-gov.net/)
The aim of this platform is to increase the transparency on the procurement procceses of public institutions and also help on better managment of the process for the procurement officials. This platform shows the data for the public contracts signed up until now throughout every step of the process.

EduPerformanca
(http://kamenica.eduperformanca.org/)
The platform was developed based on consultations with the municipalities address a very pressing issue for educational institutions in Kosovo i.e. the evaluation of the subjects taught in schools in regards to the teaching methods used, behavior, work environment etc. In turn, these evaluations that are filled by the students themselves provide crucial insight to the school lead staff as well as the municipalities to assess whether the students are satisfied with the education they are being provided, which in turn enables the identification of areas of improvement in the performance of the respective educational subject.

TREGO GJELBËR
The platform will showcase and map data of illegal dumping in Kosovo. The idea behind the platform is not just to present how the cities are unclean, but to also raise awareness about the issue and give citizens the opportunity to report cases with pictures, videos, and stories. When citizens are connected to the internet, reports of illegal dumping can be posted by using the application TrashOut. This application is linked to the platform and will automatically update the map with the information provided in the report in the pinpointed location. The platform will include information about the illegal dumping in all municipalities, so citizens, media, NGOs and other parties have a platform to turn to in order to gather information and data available regarding this matter.

WALK FREELY
(http://iwalkfreely.com/)
The app enables users the ability to report sexual harassment that they have been subjected to and provides them with data analysis tools in order to highlight trends and patterns of harassment.

8. Data Analysis

Data analysis is a process of inspecting, cleansing, transforming and modeling data with the goal of discovering useful information, informing conclusions and supporting decision-making. Data analysis is mostly used in Business, Science and Social Science domains. To analyse data the following steps need to be followed: Data Environment, data Collection, Data Processing, Data Cleaning and Data Analysis.

– Data environment includes the data that is necessary for the inputs to the analysis Example: Population, age, employment, income etc. All of these data need to be divided and have variables

– Data Collection: The data can be collected from a variety of sources for example from sensors in the environment, such as traffic cameras, satellites, recording devices, etc. Data can be obtained through interviews, downloads from online sources, or reading documentation

– Data processing: data initially obtained must be processed or organized for analysis. For instance, these may involve placing data into rows and columns in a table format (i.e., structured data)

– Data Cleaning: Once processed and organized, the data may be incomplete, contain duplicates, or contain errors. Data cleaning is the process of preventing and correcting these errors. Common tasks include record matching, identifying inaccuracy of data, overall quality of existing data, deduplication, and column segmentation

– Data Analysis: The individuals who are doing analysis with data may apply a variety of techniques to begin understanding the messages contained in the data. The techniques include Methods based on Mathematical and Statistical approaches and Machine Learning and Artificial Intelligence
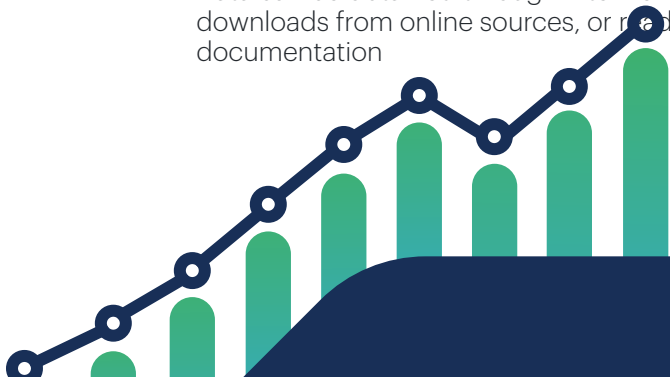
9.   Data Science

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from many structural and unstructured data. Data science is related to data mining, machine learning, and big data. Data science is a "concept to unify statistics, data analysis, machine learning, domain knowledge, and their related methods" in order to "understand and analyze actual phenomena" with data. It uses techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, domain knowledge, and information science.

10.   Data Visualization

Data Visualization is any technique of using the data for the creation of images, diagrams, or animations to communicate a message for a wider audience. Data visualization design is not about just displaying data that is collected or analyzed; it is about displaying data in a way that makes it easier to comprehend - that's where the real value lies.

Effective visualization helps users analyze and reason about data and evidence. It makes complex data more accessible, understandable and usable. Because of the way the human brain processes information, using charts or graphs to visualize large amounts of complex data is easier than poring over spreadsheets or reports. Data visualization is a quick, easy way to convey concepts in a universal manner – and you can experiment with different scenarios by making slight adjustments.
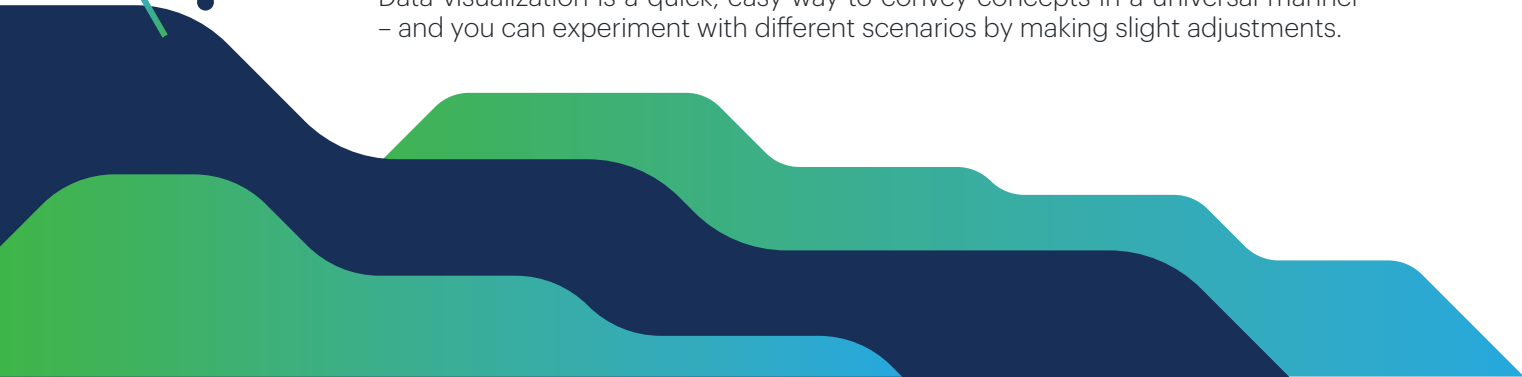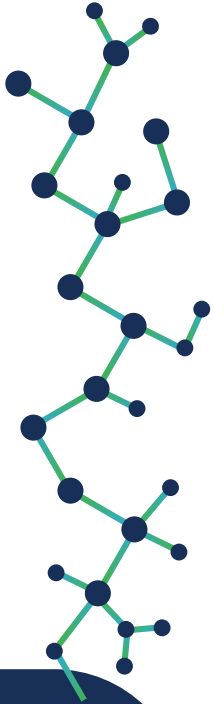
11.   Data Journalism

Data Journalism is a new set of skills for searching, understanding, and visualizing digital sources. It has a wider approach. At the core, the process builds on the growing availability of open data that is freely available online and analyzed with open source tools. Data-driven journalism strives to reach new levels of service for the public, helping the general public or specific groups or individuals to understand patterns and make decisions based on the findings. As such, data journalism might help to put journalists into a role relevant for society in a new way.

What are the story criteria?

- Relevance of the topic (Is it acute and important? Is the angle chosen originally?)

- Data Accuracy (Is your data collected and analyzed smartly?

- Are your data sources diverse and accurate?)

- Use of format (Is the format suggested relevant to your story?

- Is the story told in an innovative way?)

- Reach/Impact (how important is the story's public impact or benefit - the degree to which it serves the interests of democratic society?)

12. Open Contracting Data Standards (OCDS)

OCDS enables disclosure of data and documents at all stages of the contracting process by defining a common data model. It was created to support organizations to increase contracting transparency, and allow deeper analysis of contracting data by a wide range of users.[5]

Implementation of the Open Contracting Data Standard is an essential enhancement to an electronic government procurement system. The OCDS is a data standard that has been designed to facilitate publication and analysis of data and documents related to all stages of a contracting process. By including the OCDS, or upgrading an existing system to account for OCDS, the institution will become more transparent and also facilitate business intelligence, analysis, and monitoring that will assist policymakers and procurement practitioners to improve value for money, integrity, fairness, and performance of public contracts. Inputting all stages of procurement data through the workflow form will automatically build data aggregation that will be available to download in an open format for any user accessing the public-facing components of the platform.

It is needed that all this information when reporting to be outputted in machine-readable formats and follows by listing non-machine-readable formats as examples (Excel and PDF). Machine-readable data is data that is structured and saved in a format that can be easily processed by a computer. These can either be marked up human-readable data that can also be read by machines (e.g. microformats, RDFa, HTML) or data file formats intended principally for processing by machines (RDF, XML, JSON). File types such as Excel and PDF are human-readable but non-machine-readable formats that are made further closed by being proprietary. Use of closed proprietary formats such as Excel and PDF are actively discouraged by the international open data community in order to conform to open data best practices, particularly when hosting data in open data portals. Although we understand that the end-product should cater to user familiarity to a certain extent, one must ensure that open data standards are promoted and respected as much as possible, in particular when developing an open data portal.

5   Open Contracting Data Standards, https://www.open-contracting.org/data-standard/, accessed August 2018

13. Open Data Global Initiatives

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, and

- 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.

14.   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 taken 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 the 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 examples of Open Data in Kosovo include the 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.
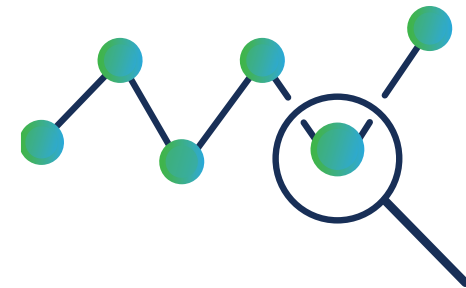
15.   About Open Data Kosovo

Open Data Kosovo (ODK) is a civic-tech NGO that recognizes the importance of technology for contributing to good governance, rule of law and social welfare. ODK believes that public data disclosure is essential in establishing effective, transparent and accountable governance. Open data is fundamental for enabling citizens to participate in decision-making; therefore, data and the information provided by government institutions should be public, easily accessible, as well as in open formats. Release of data and information held by public institutions constitutes the basic precondition for the ability of citizens, civil society and the media to utilize them for holding officials accountable and producing related positive social outcomes that result from high quality policy-making.

All local projects undertaken by ODK involve rendering 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 monitor procurement practices). Furthermore, ODK emphasizes using open data in a way that will make it possible to close the feedback loop and maximize response to the insights extracted from the data. ODK has a longstanding partnership with public institutions such as the Ministry of Public Administration, Ministry of Local Self-Governance, Municipalities, Kosovo Chamber of Commerce, etc.

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. 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 the 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 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 the openness of all public institutions in Kosovo (including thousands of indicators) and drafted policy reports[6] 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[7] 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.

—
6   https://opendatakosovo.org/publications/
7   https://ictawards.org/2017/fituesit/

Open Government Data Ecosystems