



USAID
FROM THE AMERICAN PEOPLE

Good Governance Initiative

Open Data Guidebook for Civil Servants

OCTOBER, 2020

TBILISI

The first draft of the document was prepared by the Institute for Development of Freedom of Information (IDFI) within the framework of the project “**Develop Open Data Strategies and Action Plans (2019-2020)**”. Later, the document was updated within the framework of the project “**Developing Open Data Practices in the City Halls of Gori, Lagodekhi, Ozurgeti, Senaki, Telavi and Zugdidi Municipalities**“. **The project is financially supported by the USAID Good Governance Initiative (GGI) in Georgia (USAID GGI).**

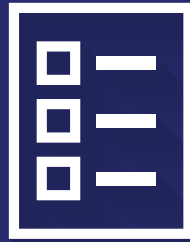
► **Disclaimer**

The views and opinions expressed in this document do not necessarily reflect the official policy or position of the United States Agency for International Development (USAID), the Good Governance Initiative (GGI), or the United States Government.

The material may not be copied, reproduced, or distributed with commercial purposes, without the prior written permission of the Institute for Development of Freedom of Information (IDFI).

Content

<i>Introduction</i>	4
<i>The Importance of Open Data</i>	8
<i>Definitions and Principles of Open Data</i>	12
<i>Open Data Definition</i>	13
<i>Open Data Principles</i>	14
<i>Convert Documents into Other Formats</i>	16
<i>Open Data Management Standards</i>	21
<i>Data Management Process</i>	22
<i>Data Processing and Publication</i>	27
<i>Practical Recommendations for Improving Open Data Management Practice</i>	39
<i>Annex #1: International Examples of Re-use of Open Data</i>	40
<i>Annex #2: Legislative Framework and Practice of Access to Public Data in Georgia</i>	45
<i>Adopting Unified Standards for Access to Open Data</i>	49
<i>Rule on Proactive Disclosure of Public Information</i>	51
<i>Unified State Registry of Information</i>	53
<i>Annex #3: Data Formats</i>	56
<i>The 5-star open data model</i>	60
<i>Annex #4: Metadata Standards</i>	67
<i>Further Reading</i>	70



Introduction

Introduction

The guidebook provides an overview of the importance, principles, and standards of open data. It also discusses the basic guidelines that civil servants should follow in the process of open government data collection, processing, and publication. The main purpose of the guidebook is to assist public institutions in overcoming data-related challenges, as well as using available resources most efficiently.¹

It is important, first of all, to distinguish **public data** from **open government data**. Each public information shouldn't be considered as open data. The format in which the information / data is processed, stored, and published is particularly important. Public information is an official document, which is stored at a public institution and is generated/created by a public agency, regardless of its format (Word, PDF document, drawing, video, photo, etc.).² Whereas, by definition, **open data is raw, also known as primary data, generated and published in a machine-readable format** (e.g., CSV, XML) that can be simply re-used by stakeholders. The guidebook summarizes the principles and standards of open data in detail.

The guidebook consists of five main parts divided into chapters. The first chapter addresses the importance of open data, followed by a definition of open data, as well as some basic principles, which should be considered when working with open data. Open data format types and characteristics are discussed in the same chapter. The next part deals with practical issues, such as the data management process and major steps from data collection till its publication. Practical tips are given for each of the stages. This chapter is followed by general tips about how to develop an effective data management system. The last part of the guidebook contains annexes, where you can find more detailed information on the following topics:

- ◆ Best international examples of applications / services developed with the use of open data;
- ◆ Legal framework of Public Data in Georgia;
- ◆ Data formats and technical features;
- ◆ Metadata standards.

¹ The list of educational materials used in developing the guidebook:

Open Knowledge Foundation. Open Data Handbook. Available at: <http://opendatahandbook.org/>

Eight Principles of Open Government Data. Available at: <https://opengovdata.org/>

European Data Portal. How to build an Open Data Strategy. Available at:

<https://www.europeandataportal.eu/en/providing-data/goldbook/how-build-open-data-strategy>

Data Opening. Good Practice Guide. Available at: <https://bit.ly/3kWLQhZ>

² Annex # 2 gives an overview of the existing legal framework regarding access to public information



The Importance of Open Data

The Importance of Open Data

Access to open government data is vital for the improvement of public service delivery and citizen participation in the decision-making process. **Open data facilitates the development of civil monitoring mechanisms, improves public well-being, and effective use of state resources.**

The public sector data is a vast source of information. According to the European Commission, direct and indirect economic benefits generated from the commercial use of data reaches up to 140 billion Euros.³ Private sector companies can use information, such as digital maps, meteorological, transport, tourist, health, financial, or economic data to create profitable business models and new services. This economic benefit is due to the availability of public data in open formats. Hence, public agencies should preferably publish and update data of public interest in the most open formats available to them, to enable stakeholders (including journalists, private sector representatives, startups, entrepreneurs, civil society members, information technology and data specialists, etc.) to use them effortlessly.

Directive 2003/98/EG (PSI Directive), adopted by the EU in 2003, highlights the importance of the availability of public data in open formats. The directive called on the EU member states to make information publicly available to the highest level possible, to ensure their reuse. The directive introduced minimum standards based on which the process of using public data for commercial purposes would be implemented in several countries. The directive aims to regulate special contractual relations between the public sector generating information and the private sector consuming it. Based on one of the recent amendments to the Directive, it is now applicable to new bodies (e.g. libraries, museums, archives). It also introduced threshold fee limits and requirements for governmental digital data.⁴

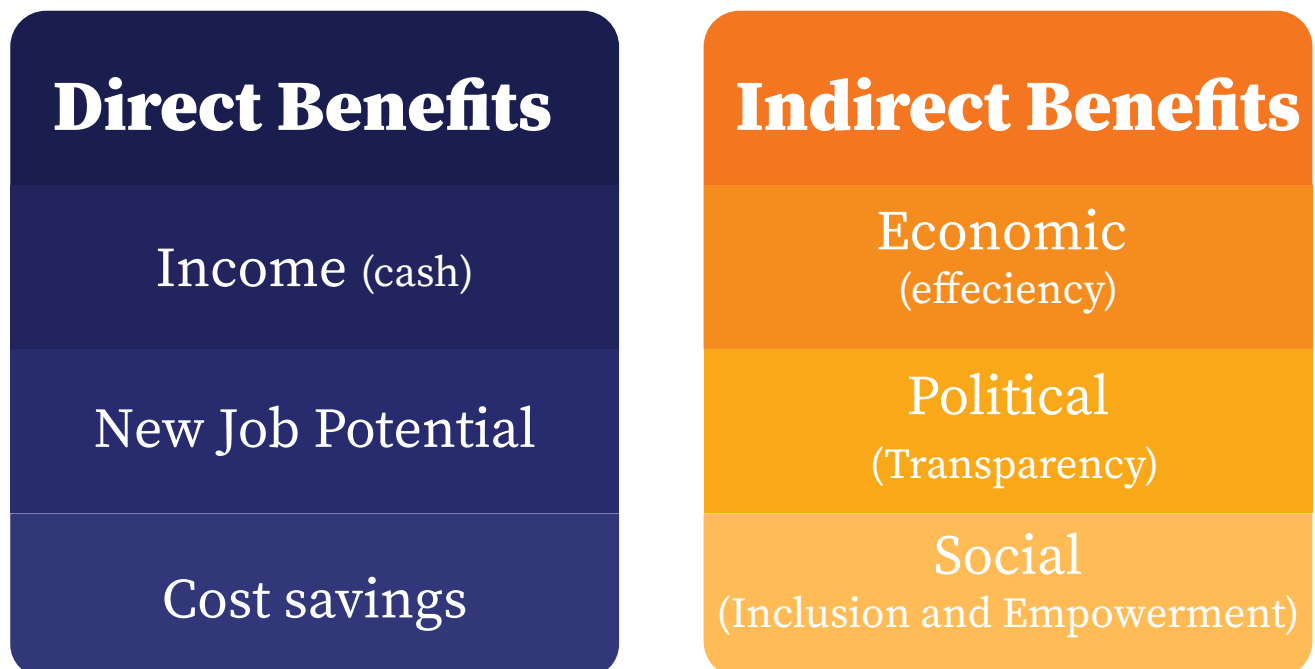
These economic aspects of transparency were exemplary for Open Government Partnership (OGP) initiative. Based on the declared principles of OGP, transparency promotes economic growth, prosperity through effective use of resources, public participation, and inclusive development. Access to relevant data and its effective utilization can serve as the basis for responsible spending of public finances, improved urban development, better educational system, comfortable public transport, etc.

Processing data in open formats first and foremost is beneficial for public agencies. The practice shows that data are re-used by public agencies, which contributes to the saving of time. Hence, **processing data in open formats enhances the data management process, reduces the possible inaccuracies as well as administrative costs.**

³ Creating Value through Open Data. European Commission. 2015. Available at: <https://bit.ly/2rRoygs>

⁴ Ibid. Available at: <https://bit.ly/2IXSIUE>

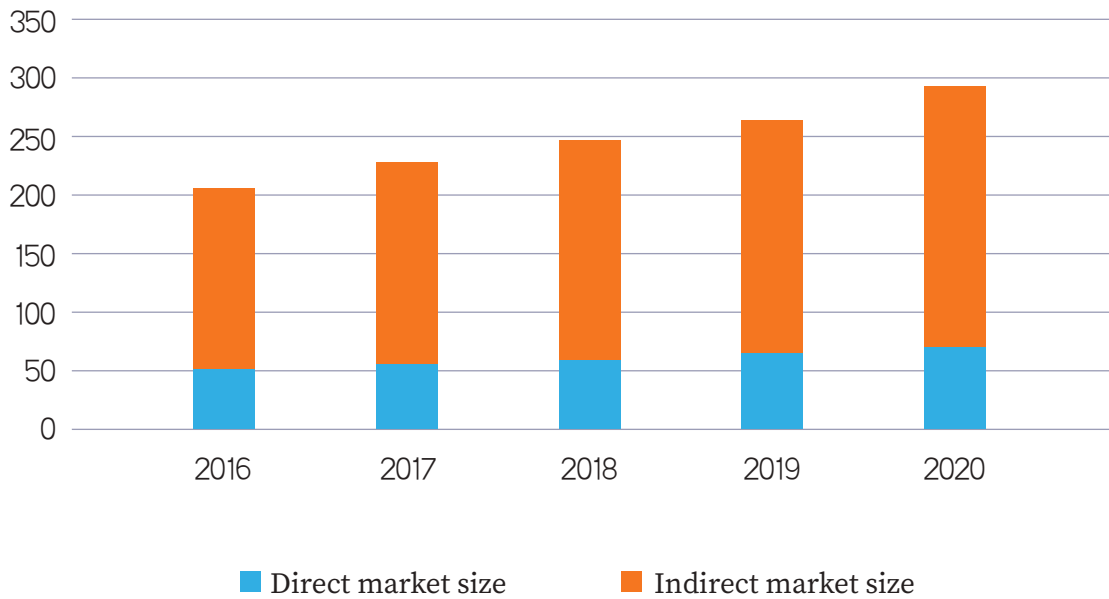
A comprehensive study was conducted in the framework of the initiative – **European Data Portal**, funded by the European Commission. This study aims to collect, assess and aggregate economic evidence to forecast the benefits of the re-use of Open Data for all 28 European Member States and the European Free Trade Association (EFTA) countries (Norway, Iceland, Liechtenstein, Switzerland), between 2016-2020. The study divides the benefit of the use of Open Data into direct and indirect benefits. Direct benefits are monetized benefits that are realized in market transactions in the form of revenues and Gross Value Added (GVA), the number of jobs involved in producing a service or product, and cost savings. Indirect economic benefits are i.e. new goods and services, time savings for the users of the applications using open data, increased efficiency, as well as transparency of the public sector, etc.⁵



Between 2016 and 2020, the market size is expected to increase by another 36.9%, from EUR 55.3 bn in 2016 to EUR 75.7 bn in 2020.

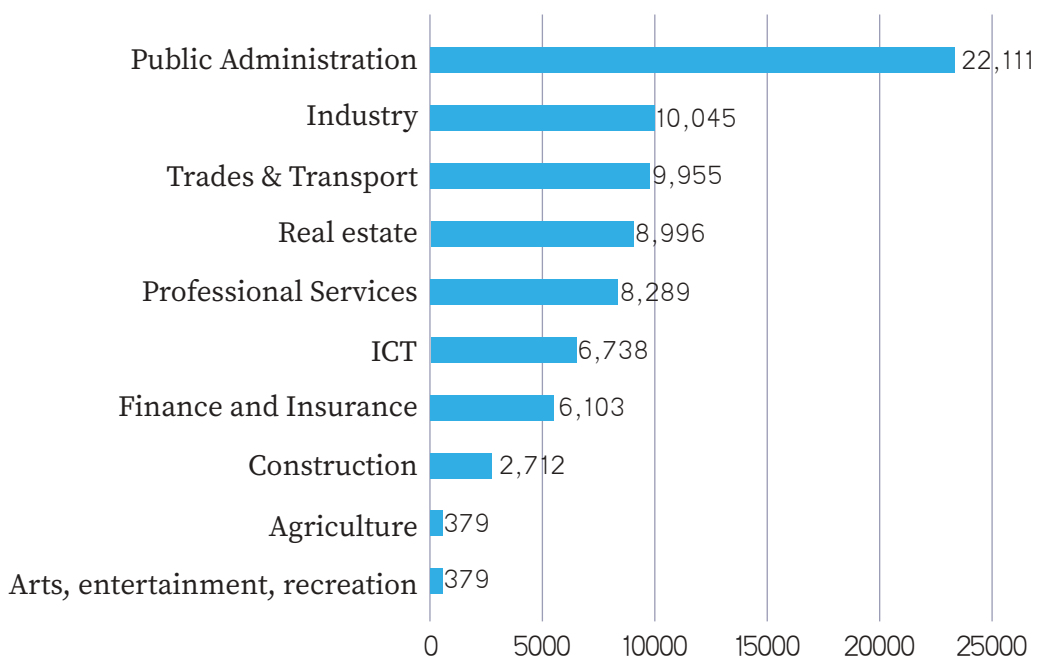
⁵ Capgemini Consulting. Creating Value through Open Data. 2015. Available at: <https://bit.ly/2IXSIUE>

Total Market size Open Data EU28+



The study also allocates the economic benefit of the use of Open Data by particular sectors. It turned out that public administration gets the greatest benefits from opening up data, followed by industry, trades, transport, and real estate. The most modest benefits are expected for the arts, entertainment, recreation, and agriculture sectors.

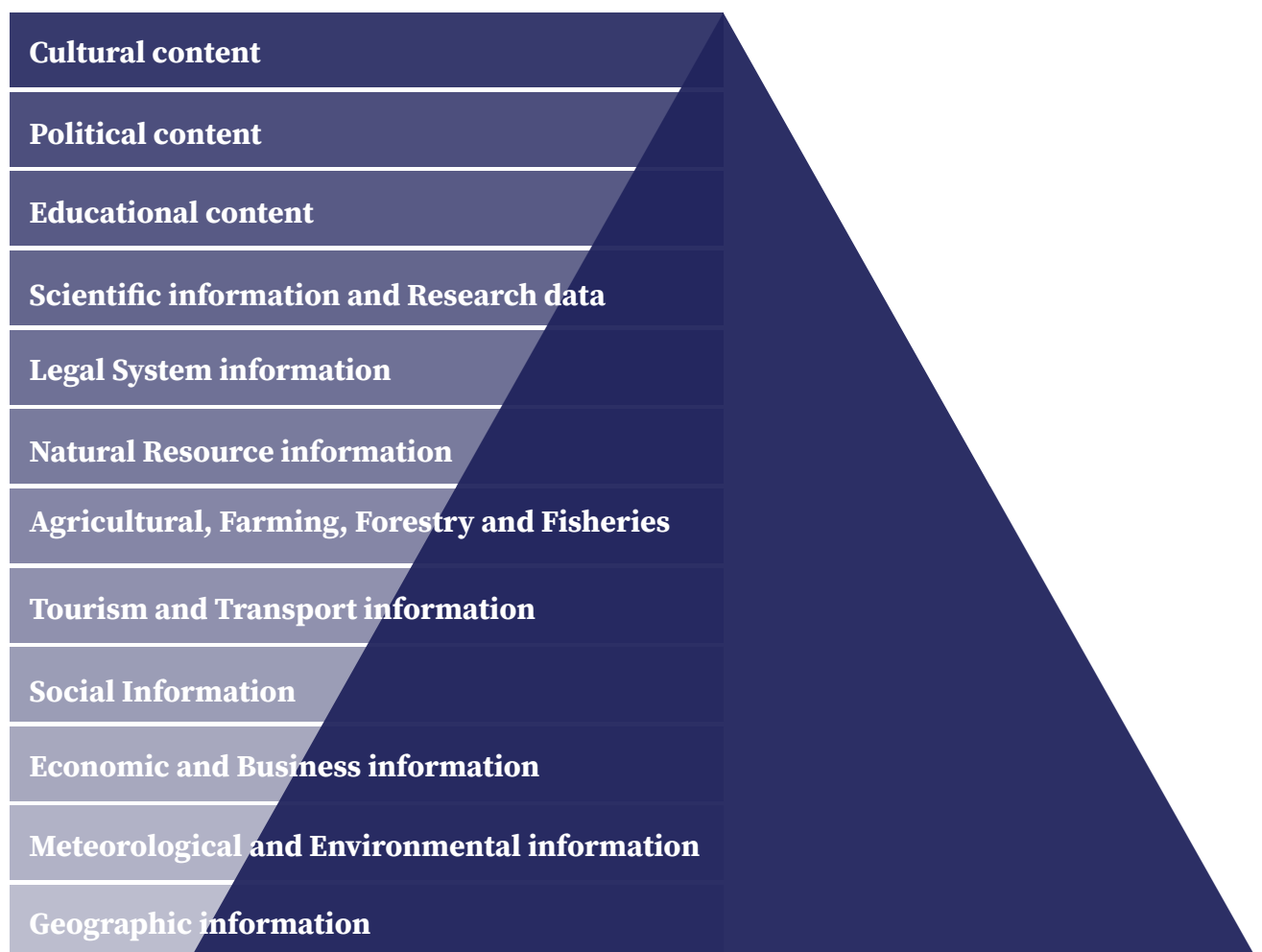
Direct Market Size per Sector (in millions)



Concerning other types of benefits, the study named the following predictions:

- ◆ The forecasted total number of direct Open Data jobs in 2016 has an upper bound of 75,000 jobs. By 2020, the upper bound provides a forecast of just under 100,000 (This equals a growth of 32% over 5 years. Per year, the increase in the number of jobs is forecasted to grow at an average rate of 7.3%). It is noteworthy that the number of jobs is not directly related to positions / services created due to the need to collect data in public agencies.
- ◆ The forecasted public sector cost savings for 2020 are EUR 1.7 bn.
- ◆ Applying Open Data in traffic can save 629 million hours of unnecessary waiting time on the roads, herewith 5.5% fewer road fatalities are expected.
- ◆ 7,000 saved lives are also expected due to quicker response to medical emergencies.

In the same study, datasets are categorized by their commercialization potential. More particular, the examples of EU countries indicate what type of data turned out to be most beneficial as a result of their reuse (see the chart below).



The domains on the bottom of the figure are recognized as having the highest commercial value. According to the chart, the domains such as geographic information, meteorological and environmental information, as well as economic, business, social, tourism, and transport contents have a high commercial potential. Although the cultural and political contents appear to have lower benefits. However, those contents are expected to concentrate on the sociocultural benefits.⁶

⁶Appendix 1 provides some examples of applications and services created using open data based on the example of different countries.



Definitions and Principles of Open Data

Definitions and Principles of Open Data

Open Data Definition

Definition of open data created by Open Knowledge Foundation: *Open data is data that can be freely used, re-used, and redistributed by anyone.*

According to this definition, some significant characteristics can be distinguished:

I. Availability - the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the Internet.

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

III. Universal Participation - everyone must be able to use, re-use, and redistribute - there should be no discriminatory restrictions.

A general definition for open data: raw, primary data, which is published as a spreadsheet in a machine-readable format (for instance: CSV, XML).

As stated before, the main distinguishing characteristics between public and open government data is the format. There are two main types of data formats:

1. Machine-readable, structured - This type of data is generated by a computer and is organized in columns and rows, whereas such data is easy to process and/or to analyze. Examples include: CSV (Comma-separated values), TSV (Tab-separated-values), Excel (.xls).

In the case of Excel (XLS), data is organized in tables that can be read using Microsoft Excel.

In CSV (comma-separated values) and TSV (tab-separated values) data is given as a direct text separated by commas and tabs. This format presents data in an encoded table, where:

- Each line is a row;
- In each line, values, separated by commas and tabs, represent columns.

2. Non-structured - sometimes data is generated by a computer but it is not organized in tables, making it typically difficult to process automatically. For example, PDF (Portable Document Format), Word, and bitmap photos (GIF, JPEG, PNG, BMP). The scanned data / documents belong to the same category, as after scanning a document, it is difficult to automatically process and analyze tables and data given in the document.

Both types of data formats have specific functions and features. See the detailed information and technical features in Annex #3.

Open Data Principles

Different sources distinguish several open data principles, including:

1.Complete - Public data should be complete and cover the topic to the fullest extent possible.

2.Primary - Open data is provided by the primary source, ensuring its trustworthiness. Alternatively, a user must be able to check whether the data is correctly aggregated. This requires metadata (general description of data), which, along with other parameters, defines how the data was created and processed.

3.Timely - Published data must be available to the public on time. It's important that compiled and processed official data is published immediately and that published data is still valuable at the time of publication.

4.Accessible - Public data must be available to the widest extent possible. Access should be simple, both in electronic and physical forms. A user should not need to visit a public institution to receive public data. For increasing accessibility of electronic data, it is recommended that such data be compiled on a unified, central platform/online location. The data should be available for download.

5.Machine-readable - Data must be structured in a way that allows its automated processing for various purposes. This principle is also called a machine-processable format.

6.Non-exclusive - The data must be available in a format that doesn't give any party exclusive rights for its distribution. It is recommended that data be available in as many different formats as possible. This will prevent anyone from creating limitations on the usage and distribution of data.

7.Non-discriminatory - The data must be available to everyone. No registration, requiring identification of a person, should be necessary.

8.Non-proprietary - The data must not be covered by any intellectual property or copyright regulations. Only limitations related to the protection and security of private data are acceptable.

9.Permanent - Access to open data should not be limited in time. Published data must remain available on a specific web address for as long as possible, allowing utilization of the data by a user with a link to the address.

When it comes to access to open data, apart from the abovementioned issues, several features should also be considered:

Available – Getting familiar with data for citizens should not be associated with additional costs. The best solution is to disseminate data online, for free.

In bulk – The data should be available as a complete set. If you have a register which is collected under the law, manage it in a way to be able to make it available as a whole for download/publication. The development of a web API or similar service may be very useful.

In an open, machine-readable format – Even more important is the provision of data in machine-readable formats, which ensures their most active users. For instance, imagine statistical data published as a PDF. While these statistics can be read by humans, but it will be difficult to process for a computer. This greatly limits the ability of others to re-use that data. Open and machine-readable formats are CSV, XML, spreadsheet. Whereas, PDF, HTML, or Text files can be re-used only after they are converted into an open format .

Using API - Data can be published via an Application Programming Interface (API). API is an interface between two applications that enables them to communicate with each other. It allows direct access to the use of the data catalogs and its functional requirements. It is possible to download the data using Request Control as it does not require any interference. They allow programmers to select specific portions of the data, rather than providing all of the data in bulk as a large file. APIs are typically connected to a database that is being updated in real-time. This means that making information available via an API can ensure that it is up to date.

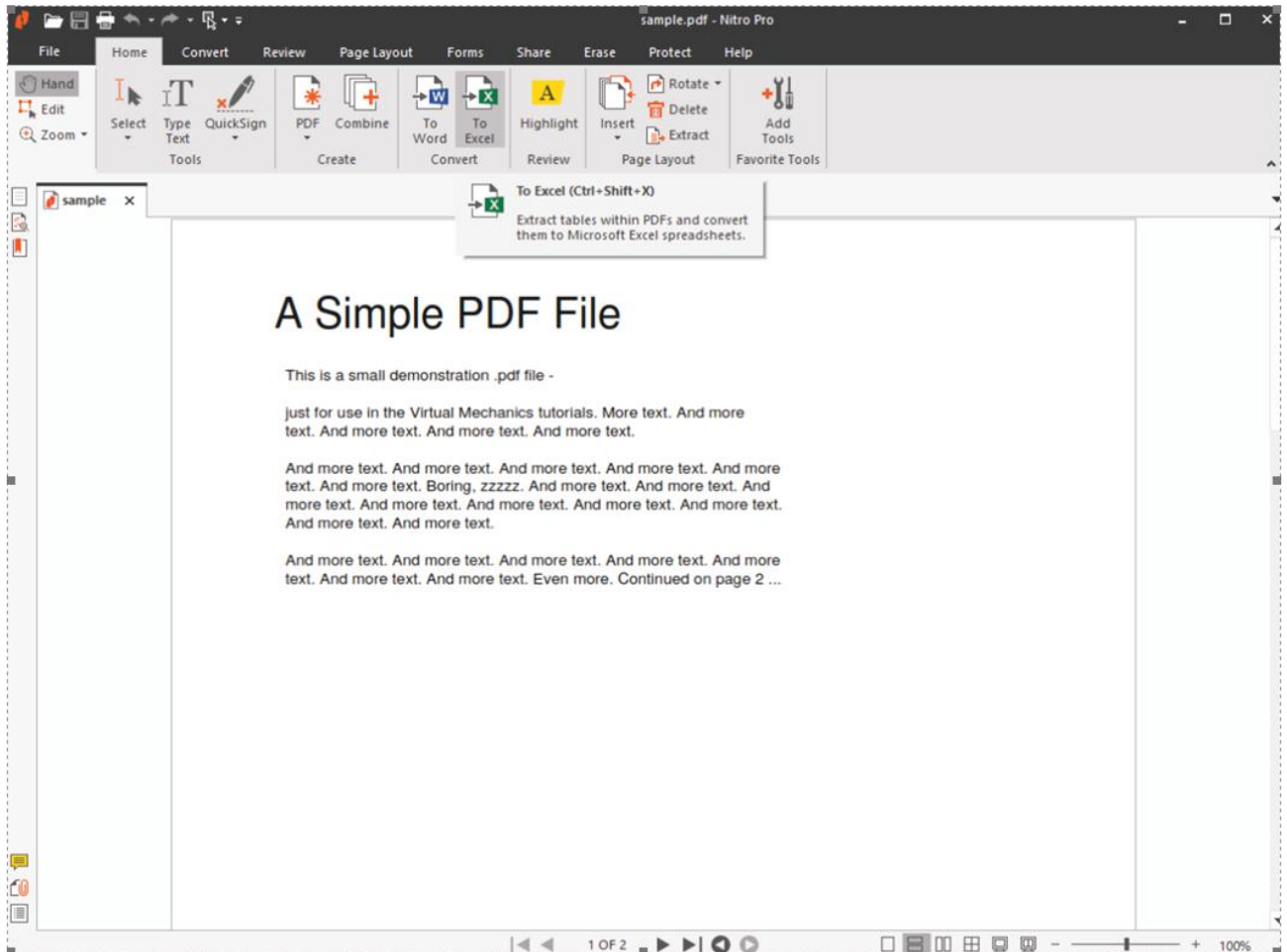
Public officials should take these principles into account when creating and implementing new standards and business processes for data collection and publication.

Convert Documents into Other Formats

Convert PDF into Excel and/or Word

First of all, attempt to convert a PDF document using a PDF reader application on the computer. Nitro PDF is recommended for this purpose.

Open your document in this application and click the button at the top left corner - to Excel. If the PDF document contains text, select **To Word**, and then the document will be converted into Word.

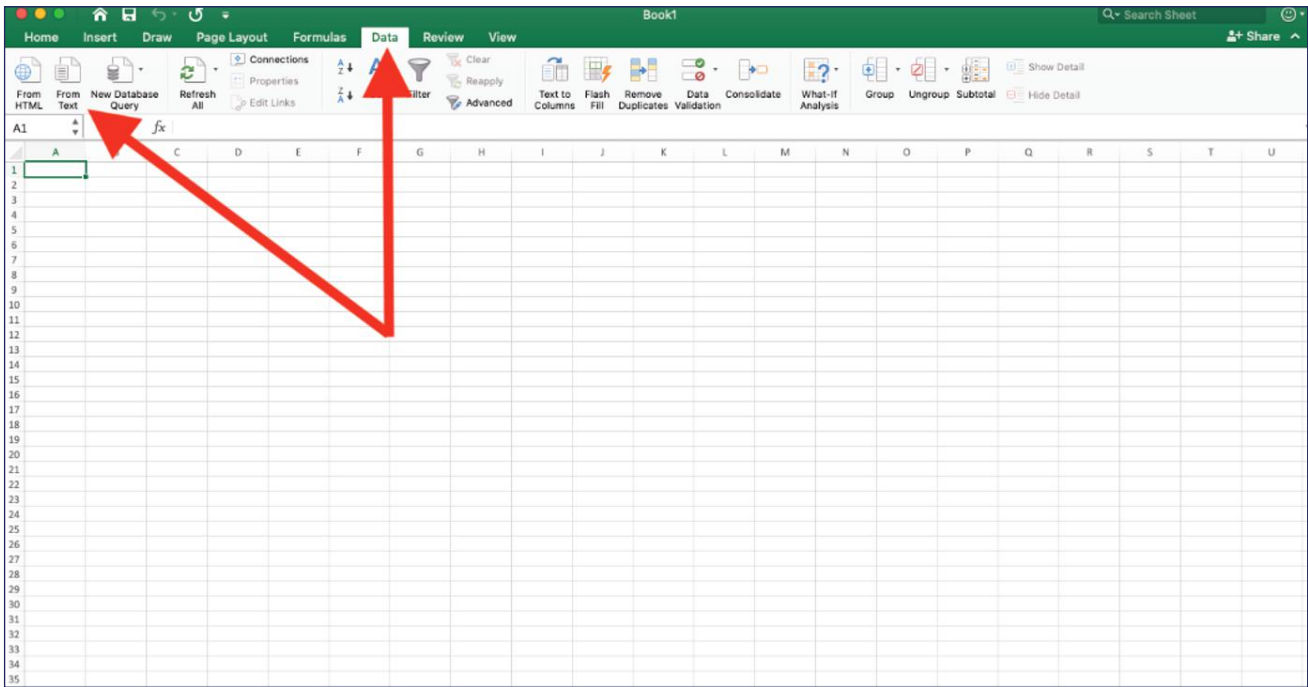


If you do not have Nitro installed on your computer, you can convert documents into different formats using the website:

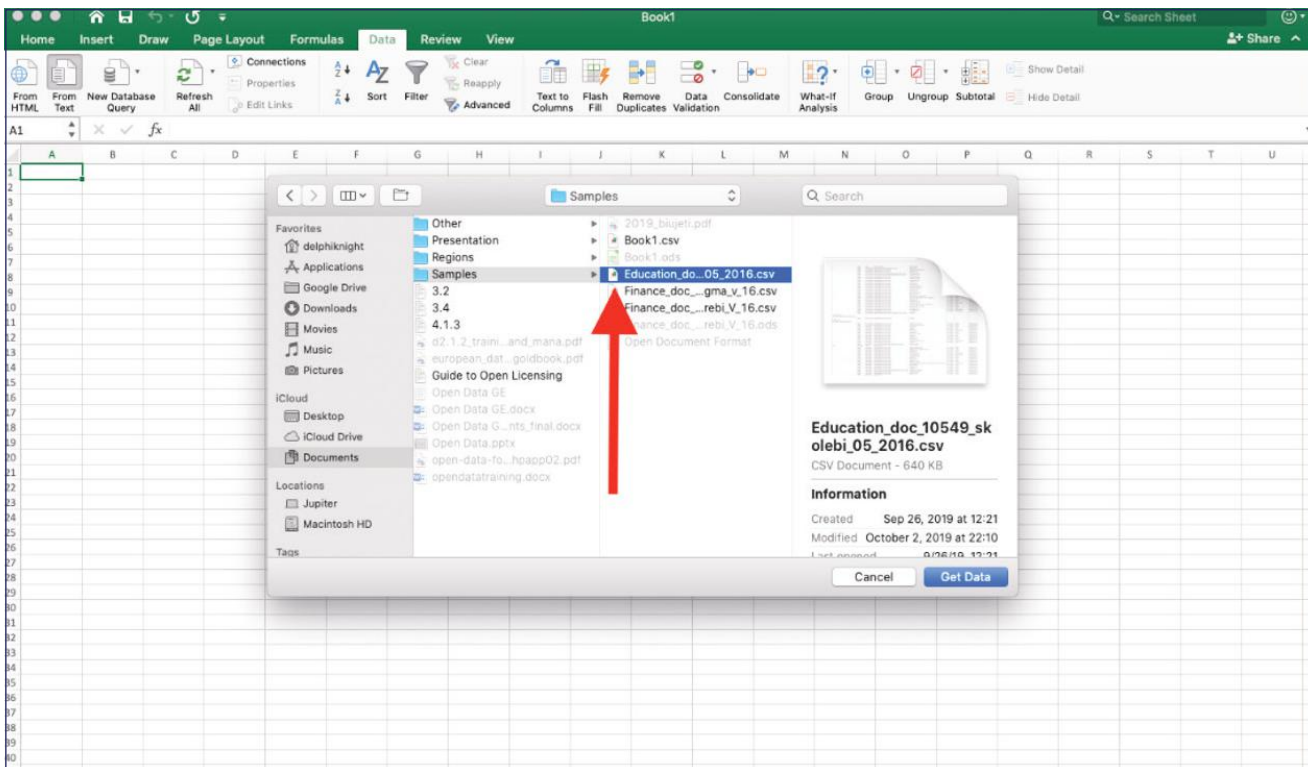
1. Visit the website: www.pdfstoexcelonline.com
2. Select from which format into which you wish to convert your document (in this case, from PDF into Excel)
3. Upload the document
4. Enter your email address where you wish to receive the document
5. Click on the link you receive in your email inbox and download the document.

Convert an Excel file into CSV

Open the Microsoft Excel worksheet, switch to the **Data** Tab. Select **Text From**, and then click on the button:



- In the **Save as** menu box, choose to save your Excel file as a CSV document.



- ◆ The Text Import Wizard will open:

Text Import Wizard - Step 1 of 3

The Text Wizard has determined that your data is Delimited.

If this is correct, choose Next, or choose the Data Type that best describes your data.

Delimited - Characters such as commas or tabs separate each field.
 Fixed width - Fields are aligned in columns with spaces between each field.

Start import at row:

File origin:

Preview of selected data:

Preview of file /Users/delphiknight/Doc.../Education_doc_10549_skolebi_05_2016.csv.

1	N	,რეგიონი	,რაიონი	,სკოლს დასახელება	,სკოლს მისამართი	,საფეხური	,სკოლს ტელეფონის ნომერი	,ელექტრონული ფოსტის მისამართი
2	1	,ობილსი	,ვაკე-საბურთალო	,ობილსის პირველი ექსპერიმენტული საჯარო სკოლა	,საირმის ქ. №84	,დენ-საბ-საშ	,2302944	,tbili
3	2	,ობილსი	,ვაკე-საბურთალო	,ქალქ ობილსის №17 საჯარო სკოლა	,აღმაშენებლის ხეივანი მე-14 კმ.	,დენ-საბ-საშ		
4	3	,ობილსი	,ვაკე-საბურთალო	,ქალქ ობილსის №35 საჯარო სკოლა	,ქავთარაძის 16/ა	,დენ-საბ-საშ	,2302944	,tbili
5	4	,ობილსი	,ვაკე-საბურთალო	,მიხეილ გრუშვეცკის სახელობის ქალქ ობილსის №41 საჯარო სკოლა	,ზ. ფლაშვილის ქ.			
6	5	,ობილსი	,ვაკე-საბურთალო	,ქარაფელი იუნკერი სახელობის ქალქ ობილსის №52 საჯარო სკოლა	,აღმაშენებლის ხეივანი მე-14 კმ.	,დენ-საბ-საშ		
7	6	,ობილსი	,ვაკე-საბურთალო	,ქალქ ობილსის №54 საჯარო სკოლა	,ზ. ფლაშვილის ქ. №59	,დენ-საბ-საშ	,2221156	
8	7	,ობილსი	,ვაკე-საბურთალო	,ქალქ ობილსის №55 საჯარო სკოლა	,ი. ჭავჭავაძის გამზ №38	,დენ-საბ-საშ	,2292472	
9	8	,ობილსი	,ვაკე-საბურთალო	,ქალქ ობილსის №60 საჯარო სკოლა	,აღმაშენებლის ხეივანი მე-14 კმ.	,დენ-საბ-საშ	,2397429	

- ◆ Make sure that the Delimited radio button and Unicode (UTF-8) in the File origin list are selected and then click the button Next >

Text Import Wizard - Step 2 of 3

This screen lets you set the delimiters your data contains.

Delimiters

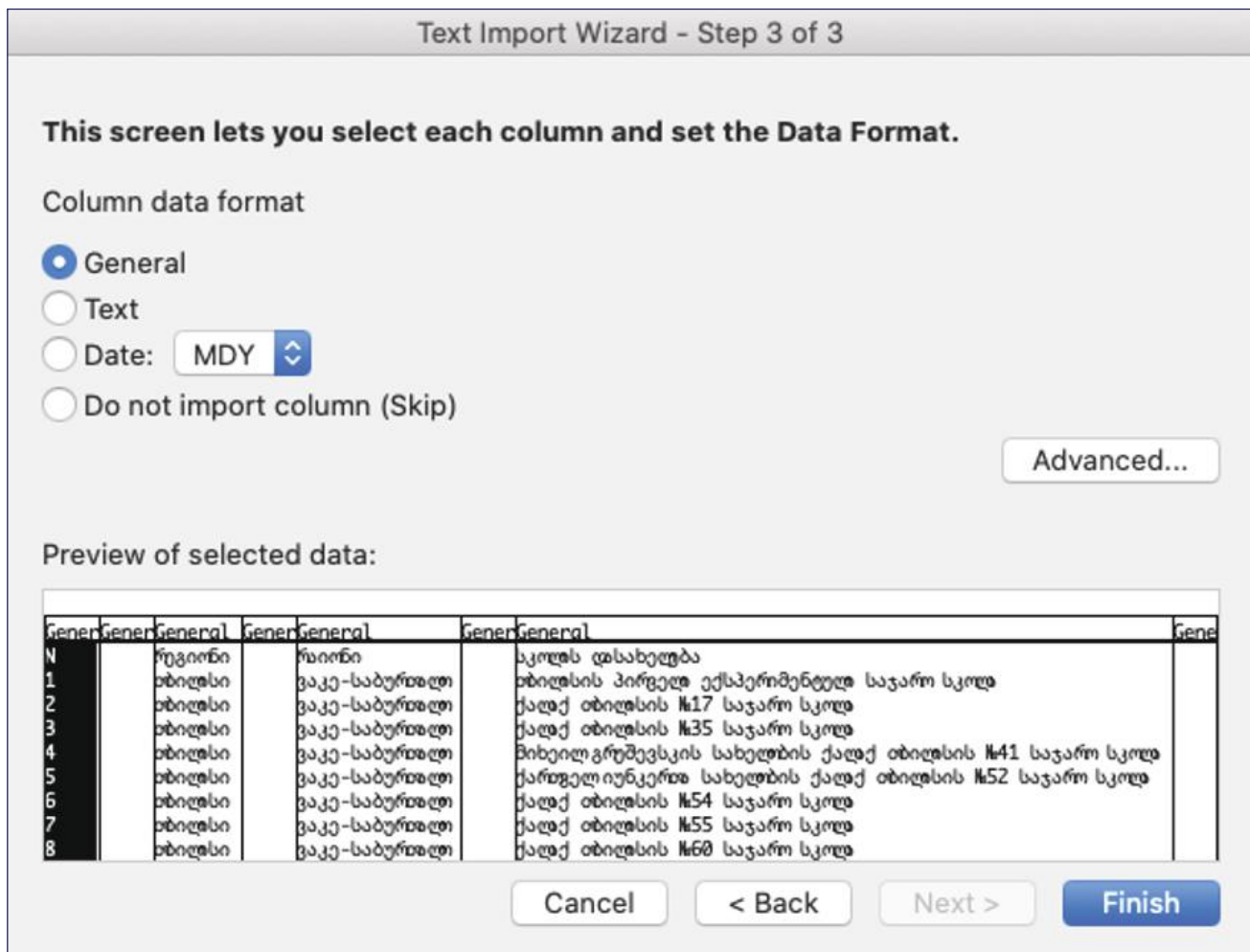
Tab
 Semicolon
 Comma
 Space
 Other:

Treat consecutive delimiters as one
 Text qualifier:

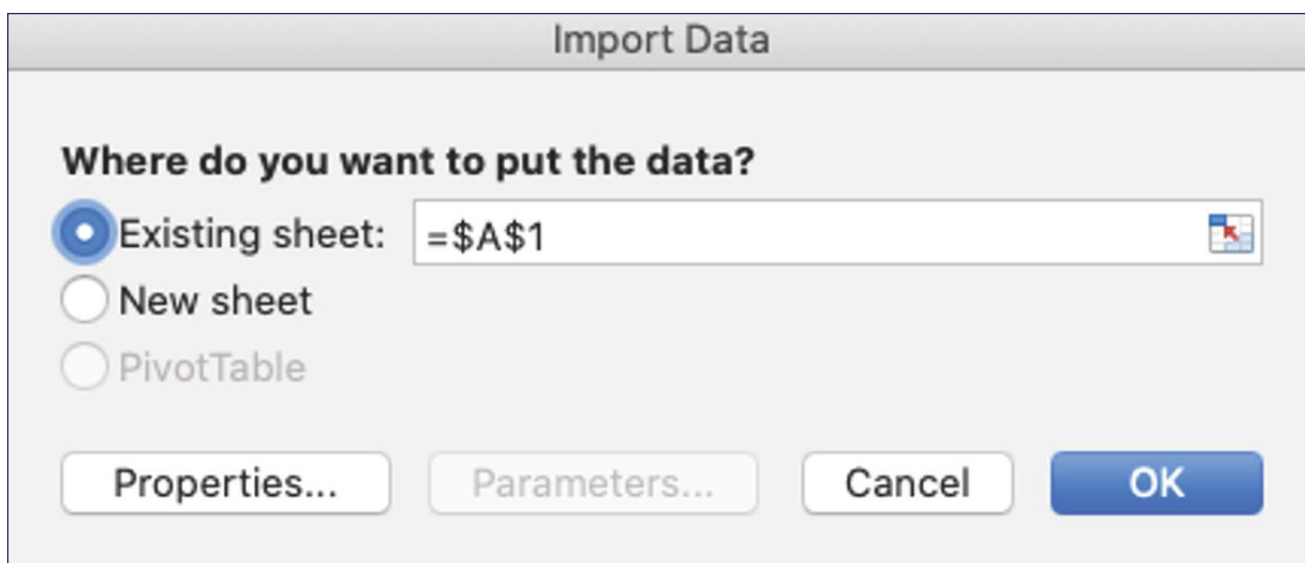
Preview of selected data:

N		რეგიონი	რაიონი	სკოლს დასახელება	სკოლს მისამართი	საფეხური	სკოლს ტელეფონის ნომერი	ელექტრონული ფოსტის მისამართი
1		ობილსი	ვაკე-საბურთალო	ობილსის პირველი ექსპერიმენტული საჯარო სკოლა	საირმის ქ. №84	დენ-საბ-საშ	2302944	tbili
2		ობილსი	ვაკე-საბურთალო	ქალქ ობილსის №17 საჯარო სკოლა	აღმაშენებლის ხეივანი მე-14 კმ.	დენ-საბ-საშ		
3		ობილსი	ვაკე-საბურთალო	ქალქ ობილსის №35 საჯარო სკოლა	ქავთარაძის 16/ა	დენ-საბ-საშ	2302944	tbili
4		ობილსი	ვაკე-საბურთალო	მიხეილ გრუშვეცკის სახელობის ქალქ ობილსის №41 საჯარო სკოლა	ზ. ფლაშვილის ქ.			
5		ობილსი	ვაკე-საბურთალო	ქარაფელი იუნკერი სახელობის ქალქ ობილსის №52 საჯარო სკოლა	აღმაშენებლის ხეივანი მე-14 კმ.	დენ-საბ-საშ		
6		ობილსი	ვაკე-საბურთალო	ქალქ ობილსის №54 საჯარო სკოლა	ზ. ფლაშვილის ქ. №59	დენ-საბ-საშ	2221156	
7		ობილსი	ვაკე-საბურთალო	ქალქ ობილსის №55 საჯარო სკოლა	ი. ჭავჭავაძის გამზ №38	დენ-საბ-საშ	2292472	
8		ობილსი	ვაკე-საბურთალო	ქალქ ობილსის №60 საჯარო სკოლა	აღმაშენებლის ხეივანი მე-14 კმ.	დენ-საბ-საშ	2397429	

- ◆ Select Comma as a column delimiter symbol, and then click the button Next >



- ◆ Use the column formats section to select the format for each column. Although in most cases, it is not required. Click on the **Finish** button.



- ◆ You might be asked from which cell to start importing data. The import starts from the upper-left corner (A1). Click 'OK' to proceed.

A	B	C	D	E	F	G	H	I
1	N	რეგიონი		რაიონი		სკოლის დასახელება		სკოლის მისამართი
2	1	თბილისი		ვაკე საბურთალო		თბილისის პირველი ექსპერიმენტული საჯარო სკოლა		საბურთალოს რაიონის ქ. №84
3	2	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №17 საჯარო სკოლა		აღმაშენებლის ხეივანი მე-14 კმ.
4	3	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №35 საჯარო სკოლა		ქავთარაძის 16/ა
5	4	თბილისი		ვაკე საბურთალო		მხივილ გრემუჯვის სახელობის ქალაქ თბილისის №41 საჯარო სკოლა		ზ. ფალიაშვილის ქ. №34
6	5	თბილისი		ვაკე საბურთალო		ქართველი იუნკერთა სახელობის ქალაქ თბილისის №52 საჯარო სკოლა		აღ. ყაზბეგის გამზ. №27
7	6	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №54 საჯარო სკოლა		ზ. ფალიაშვილის ქ. №59
8	7	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №55 საჯარო სკოლა		ი. შავჭავჭავაძის გამზ. №38
9	8	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №60 საჯარო სკოლა		აღ. ყაზბეგის გამზ. №39
10	9	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №61 საჯარო სკოლა		დოლიძის ქ. №3
11	10	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №62 საჯარო სკოლა		მარტყვაძის ქ. №20
12	11	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №64 საჯარო სკოლა		ვაკა-ფშაველას გამზ. №6-ა
13	12	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №102 საჯარო სკოლა		მუცხეშვილის პლატო, მე-4 მ/რ
14	13	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №122 საჯარო სკოლა		ქავთარაძის ქ. №23ა
15	14	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №123 საჯარო სკოლა		იმედაშვილის ქ. №4
16	15	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №126 საჯარო სკოლა		ვაკა-ფშაველას გამზ. №71
17	16	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №128 საჯარო სკოლა		თოფურის ქ. №12
18	17	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №151 საჯარო სკოლა		ვაკა-ფშაველას №61, მეორე კვარტალი
19	18	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №158 საჯარო სკოლა		ვახუშტის რაიონის ქ. მეორე შესახვევი №21
20	19	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №161 საჯარო სკოლა		ოსიპიძის ქ. №50
21	20	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №162 საჯარო სკოლა		ფანასკერტელის ქ. №12
22	21	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №165 საჯარო სკოლა		ზ. ფალიაშვილის ქ. №57
23	22	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №171 საჯარო სკოლა		მუცხეშვილის პლატო, III მ/რ II კვ. ძონგინძის ქ. №3
24	23	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №186 საჯარო სკოლა		დიდი დოლიძის, პეტრინის ქ. №8ა
25	24	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №192 საჯარო სკოლა		დიდი დოლიძის III მ/რ
26	25	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №195 საჯარო სკოლა		ს. ჩიქოვანის ქ. 26
27	26	თბილისი		ვაკე საბურთალო		ქალაქ თბილისის №198 საჯარო სკოლა		მარტყვაძის ქ. №20
28	27	თბილისი		ვაკე საბურთალო		ვლადიმერ კომაროვის თბილისის ფილიალ - მათემატიკის № 199 საჯარო სკოლა		ქ. თბილისი
29	31	თბილისი		რსანი-სამგორი		ქალაქ თბილისის №65 საჯარო სკოლა		გორგასლის №75
30	32	თბილისი		რსანი-სამგორი		ქალაქ თბილისის №70 საჯარო სკოლა		ფორციას ქ. №3
31	33	თბილისი		რსანი-სამგორი		ქალაქ თბილისის №72 საჯარო სკოლა		გორგასლის ქ. №2
32	34	თბილისი		რსანი-სამგორი		ქალაქ თბილისის №74 საჯარო სკოლა		კაკაბეთის ქ. №2
33	35	თბილისი		რსანი-სამგორი		ქალაქ თბილისის №76 საჯარო სკოლა		აეროპორტის დასახლება
34	36	თბილისი		რსანი-სამგორი		ქალაქ თბილისის №81 საჯარო სკოლა		ჩკალაძის №3
35	37	თბილისი		რსანი-სამგორი		ქალაქ თბილისის №82 საჯარო სკოლა		ნაკაძის ქ. №6
36	38	თბილისი		რსანი-სამგორი		ქალაქ თბილისის №83 საჯარო სკოლა		მოსკოვის გამზ. №37
37	39	თბილისი		რსანი-სამგორი		ქალაქ თბილისის №85 საჯარო სკოლა		ბერი გაბრიელ სალოსის გამზ. №131
38	40	თბილისი		რსანი-სამგორი		ქალაქ თბილისის №86 საჯარო სკოლა		ბელეთის №12
39	41	თბილისი		რსანი-სამგორი		ვაკა-ფშაველას სახელობის ქალაქ თბილისის №87 საჯარო სკოლა		ქეთევან ნაძებულის
40	ქ. №90/2 *	დან-საბ-საბ		2775133		tblilisi87@mes.gov.ge		
41	47	თბილისი		რსანი-სამგორი		რსანი-სამგორის №88 საჯარო სკოლა		ჩიხრაძის დასახლების ქ. №40

- ◆ Once confirmed, your CSV document will be converted into an Excel document.



Open Data Management Standards

Open Data Management Standards

Data Management Process

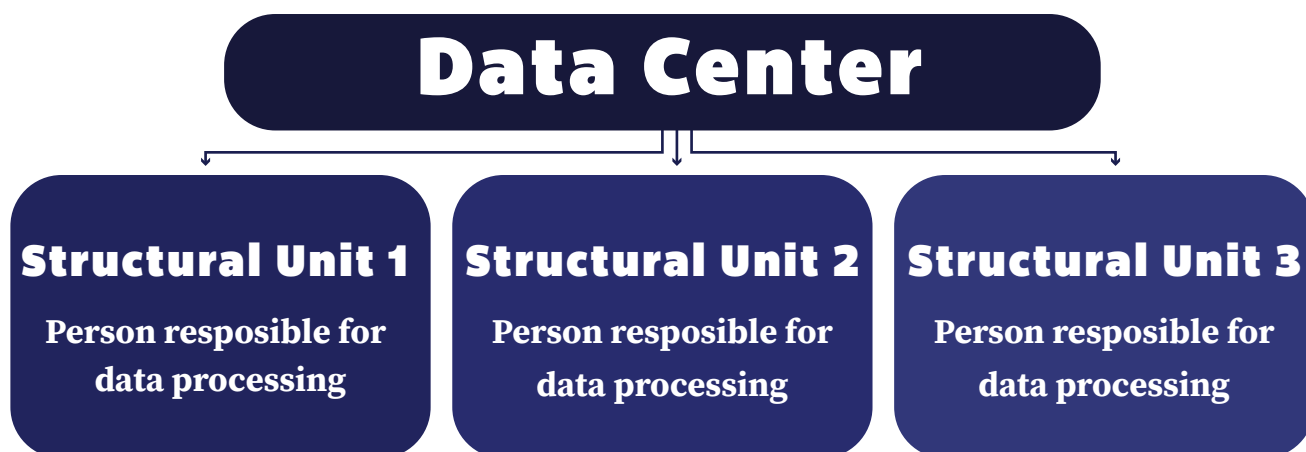
Any public institution should have a data management system, which implies having a unified data management process. This process can be described in a special instruction, rule, or internal normative act. To develop effective internal procedures for data management, it is important to consider the following circumstances:

- ◆ Unified data management process should be comprehensive and cover all the structural units collecting particular public data;
- ◆ This process should be overly bureaucratic / hierarchical, and should be as flexible as possible;
- ◆ Public servants should be aware of the policies and legislation regarding public data;
- ◆ People responsible for data management should recognize the benefits of publishing data in open formats;
- ◆ Different structural units must cooperate with each other in creating and sharing comprehensive datasets.

Public agencies of different sizes may have various data management structures. A public institution should create a structure relevant for its work, size, and goals. To create an effective structure / system and properly implement it, consider some circumstances:

- ◆ The importance of producing and publishing data in open formats must be recognized at the political level, which will ensure that the issue is a priority and there is a political will.
- ◆ In the process of data management, the leader / responsible person of the process should be identified and his / her relevant rights and responsibilities should be described.
- ◆ Public servants responsible for and engaged in the data management process should also be determined at the level of structural units. They should assist the team leader in achieving the objectives set out in regard to data openness.
- ◆ There should exist a data openness policy / strategic vision, which includes data processing-publishing standards and the person(s) responsible for its implementation, frequency of data publication, unified list / catalog of data stored in a public institution, etc.

Based on the particularities of the municipal bodies, it is appropriate to consider the following structure:



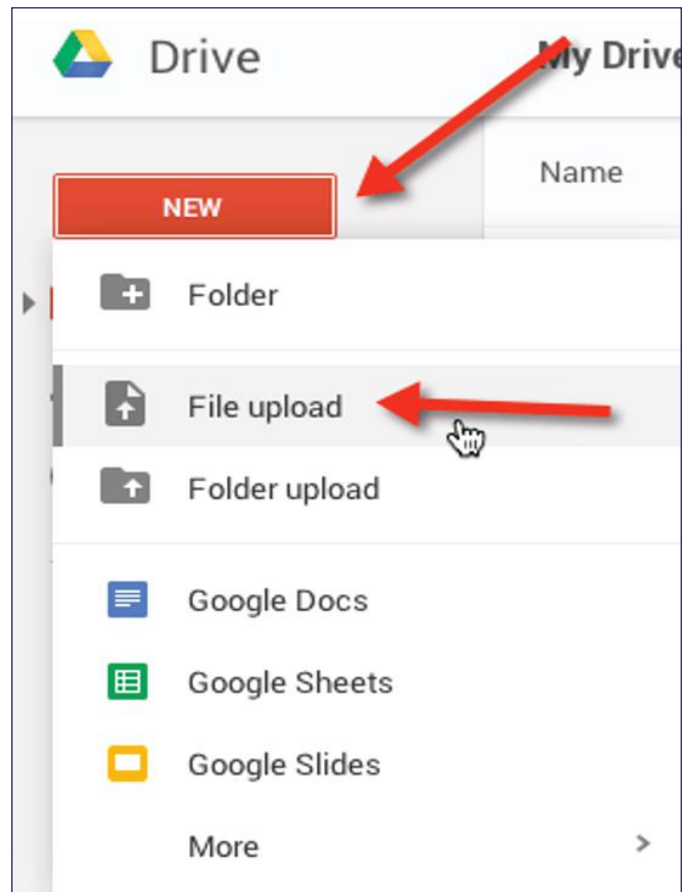
According to this structure, each structural unit is responsible for collecting and preparing data. However, the data they collect are accumulated in one place. In the case of such a structure, there is no necessity to set up an additional structural unit, where the data, stored in the City Hall, will be gathered. Instead, it is always recommended to identify the structural unit, under which the data is stored centrally and, if necessary, easily exchanged with other structural units. Furthermore, data may be stored in a shared folder on the municipality's internal network (i.e. an electronic, centralized data system), which will be accessible to the relevant staff of each unit. Such practice facilitates coordinated data processing and simplifies joint work on particular datasets among employees.

A centralized data platform / data center can be both, a physical server and a virtual platform.

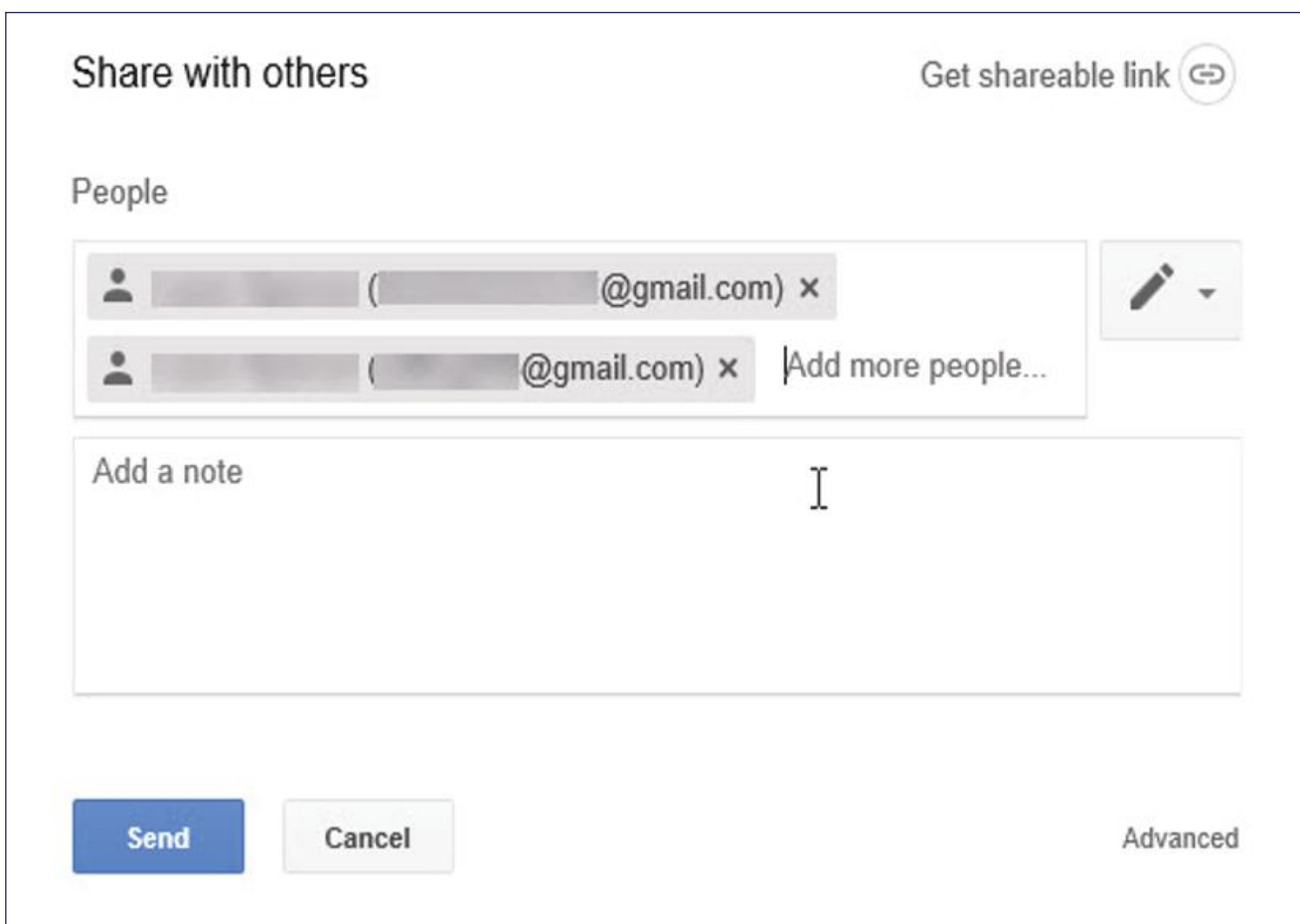
In the case of a physical server, the centralized database is located in a single computer that can be connected to a local network. Therefore, other computers may be connected to that database. To this end, an internal network should be set up. Any authorized user of a public institution will have an access to the server, while the administrator will have additional technical functions, including changing system configuration and determining what type of file a particular group / person has access to, whether the user can view the file, add a new one, delete it and / or edit it.

When it comes to cloud storage/online data storage management tools, **Google Drive**⁷ can be considered as one of the safest file repositories that allow any Google user to create or upload an existing file/folder, photo, video, or any other material, providing flexible storage options. The latter enables Google users to store and access files online. To use Google Drive, you need a Gmail account.

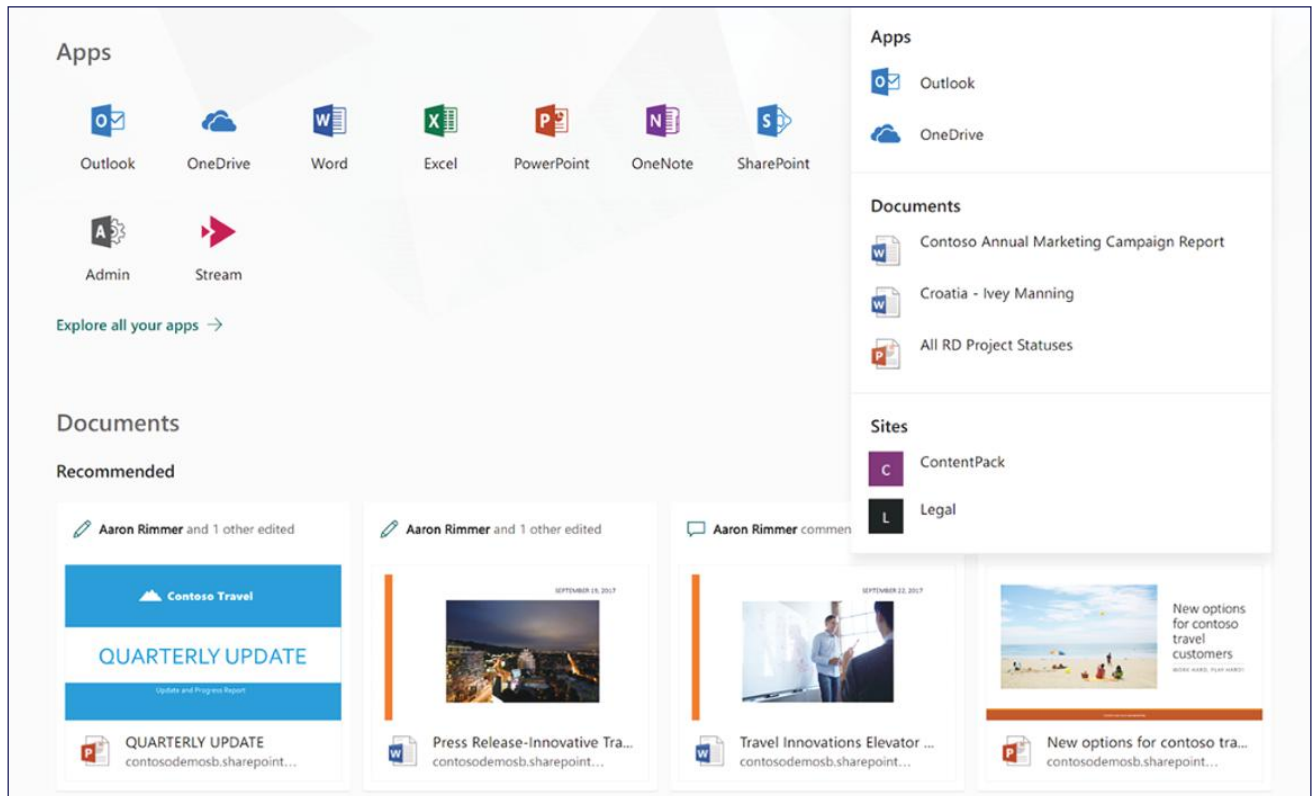
⁷ <https://www.google.com/drive/>



Google Drive enables you to process and save text / digital information (like Microsoft Word and Excel). Also, Drive has a sharing feature. Several people can access and edit the material at the same time, which simplifies team work. It is also possible to restore and / or revise previous, revised versions of the document.



Microsoft Office 365⁸ can be considered as an alternative option to Google Drive with its technical features. For both platforms, data is stored in the cloud storage, but Office 365 also has some advantages, including the abundance of Excel chart types, the functional diversity of Word, the simplicity of the calendar, and more. The Microsoft Office 365 account is paid, and the monthly price varies by service and supported application.



A team leader / person responsible for data takes a leading role in coordinating the work of the structural unit and is also responsible for the implementation of the open data policy. Furthermore, a team leader has information about the needs of citizens, representatives of civil society organizations, businesses, and other stakeholders. To succeed, the person occupying this position must consider the following circumstances:

- ◆ He/she should have **a vision** about the potential benefits of particular data (based on specific examples) and especially their re-use.
- ◆ He/she should explain/show the advantage of providing data in open formats to the heads of particular structural units of a public institution.
- ◆ He/she should provide **capacity building opportunities to the public servants** engaged in data management and publication.
- ◆ **It is important to establish active communication with all stakeholders.** Even though they are not involved in data processing daily, they are the actors using data in everyday work. Thus, it is crucial to get their feedback in identifying, publishing, and updating new datasets.

⁸ <https://www.office.com/>

- ◆ **It is of great importance to establish partner relationships with different stakeholders**(representatives of business, civil society, and think tanks). Applying inclusive approaches can rebuild citizen trust which may lay the foundation for the development of open data initiatives and services through public-private partnerships.

As for the **data management team**, composed of persons responsible for data management from particular structural units, their responsibilities include:

- ◆ Develop, coordinate, monitor, and evaluate an internal data management/ publication policy;
- ◆ Identify and develop open data standards;
- ◆ Register and analyze datasets stored in the public institution;
- ◆ Identify and develop software necessary for data processing and publication;
- ◆ Draw up a data publication schedule;
- ◆ Cooperate with stakeholders interested in data publication and use.



Data Processing and Publication

Data Processing and Publication

Data processing and publication consist of several stages, including collection, preparation, publication, and storage. In this section, the document will analyze each stage of data processing and publication as well as the circumstances a public servant should consider during these stages.



1. Data Collection

The data collection method is particularly important as the datasets to be produced / processed and later to be published by a public institution are identified at this stage. The formats are also determined/planned at this stage. During this process, it is of great importance to consider the interests of various stakeholders to ensure the effective use of the data in the future.

Additionally, the data collection stage includes several sub-phases, such as identifying existing data, identifying priority datasets, describing datasets, identifying the data for publication, and categorizing datasets.



During data collection, the public institutions should accumulate all datasets at their disposal. Datasets should be identified, and then the most relevant/priority datasets should be determined. When evaluating each dataset, representatives of a public institution should consider the following circumstances:

- ◆ Whether the public institution has relevant resources for data production and dissemination (in terms of political, legal, or organizational dimensions)?
How complete and accurate is the dataset?
- ◆ Does the data need to be cleaned?
- ◆ What is the format of the data?
- ◆ Is data of high public interest?
- ◆ What are the possible benefits of re-using data?

While identifying existing datasets, **a public institution can collect/obtain data in a variety of ways:**

- ◆ Gather datasets created/produced by different units and group them based on common categories;
- ◆ Use existing electronic programs and extract interesting data from them. For example, most municipalities have a State Treasury Electronic Service System, set up by the Ministry of Finance of Georgia, as well as electronic budget and human resource management systems from which the data of public interest can be extracted and published;
- ◆ Cooperate with other public institutions and obtain data relevant to its activities;
- ◆ Collaborate with stakeholders and conduct individual interviews with individuals connected with datasets;
- ◆ Conduct surveys on particular issues and keep the results of the survey as raw, primary data;
- ◆ Maintain detailed datasets of beneficiaries of public services by keeping detailed information about beneficiaries;
- ◆ Register real-estate and/or assets owned by a public institution and their characteristics;
- ◆ Record/register the contracts signed by the public institution and particular details of the contract (parties, subject of the contract, contract value, terms, etc.);
- ◆ Conduct a population census;
- ◆ In the case of cities and self-governing units, record important facilities, buildings, institutions, etc. within the city/municipality, etc.

In addition to identifying datasets stored in a public institution, **it is important to identify the types of data that are not systematically collected yet**, but the production of which might be useful. Such datasets can be identified in different ways:

- ◆ Identify information that is often requested by various stakeholders through FOI requests;
- ◆ Consult with various stakeholders to identify the data needed for developing new services/projects;
- ◆ Start collecting and producing relevant data in response to new local challenges, policies, or regulations;
- ◆ Producing detailed datasets of services provided by the legal entities under the municipality (the legal entities of public/private law owned or governed by municipalities), including the beneficiaries of these services;

- ◆ Collaborate with various local and national stakeholders, including representatives of the international community regarding data collection issues. The production of particular datasets can be considered as one of the components of the projects implemented in cooperation with these actors in different directions.

When producing new datasets, it is important data are collected and grouped based on a unified standard. At this point, it is crucial to collect data with the use of as detailed categories and criteria as possible (e.g. year, age, first name, last name, public institution, region, etc.). The more detailed data categories are, the more comprehensive analysis can be conducted based on these data in the future. Therefore, the following circumstances must be taken into account when processing datasets:

- ◆ All employees working on the dataset should have a deep understanding of the unified data management standard and of all categories of data stored in the dataset;
- ◆ Employees working on the dataset should strictly follow the unified data management standard.

Alongside identifying and collecting priority datasets, metadata i.e. general description of a dataset, should also be created for each dataset. In particular, the following general details should be kept, which will be later published together with a dataset:

- ◆ Title and type of data;
- ◆ Data publication and latest update dates;
- ◆ Legal issues - whether personal data protection is required;
- ◆ Technical issues – for example, publication format, license;
- ◆ Organizational issues – for example, public institution/unit, which processed/published data, frequency of data updates, etc.
- ◆ To what extent it is possible to publish data in an open format.

The data that will be published first and foremost should be determined with the consideration of the activities of a public institution and the public interest. For instance, it is widely acknowledged that the most valuable open government datasets are the registry of companies and businesses, crime statistics, meteorological data, information on agricultural land, hunting, fisheries and forestry, data on education, energy, environment, health, and social policy, as well as, financial data and procurement, digital maps, administrative costs, legal acts, demographic data, transport, and infrastructure. Datasets of public interest can also be identified by analyzing public information requested over the past years and months and by taking into account the opinions of stakeholders (e.g., private and civil society, media, data specialists, and researchers).

After identifying the most valuable datasets, their management and categorization issues should be planned. In particular, datasets should be grouped according to the structural units and/or main topics. At the same time, a team of employees should be determined responsible for data processing, update, and accuracy.

2. Data Preparation / Processing

After accumulating all datasets stored in a public agency, it is important to process, organize, and ensure data accuracy, which is a prerequisite for preparing data for publication. Four issues should be considered while their publication: quality, technical openness, legal openness, and metadata.

As mentioned above, data should be collected and updated based on a predetermined standard. At the same time, the following circumstances should be taken into account:

- ◆ Any dataset should be updated with a predefined frequency (unless there is an exception).
- ◆ It is preferable to introduce a data cross-checking practice, under which data will be double-checked and approved with a predetermined frequency. This approach will ensure data accuracy.
- ◆ Apart from the data update, you may need to use particular functions of Excel, to extract relevant information from the primary / raw data for analysis, publication and / or provision of public information. To this end, the following function is useful:
 - Sum
 - Percentage change
 - Subtracting
 - Average and Median
 - COUNTA function - calculate the number of cells in a column
 - COUNTBLANK function - calculate the number of blank cells
 - COUNTIF function – calculate the values from a large dataset based on particular parameters
 - COUNTIFS function - group and calculate data according to particular criteria (e.g. age group, region, etc.)
 - Sorting - sort data according to particular criteria
 - Excel Pivot Tables - process raw data and group them based on various criteria.⁹

⁹ See additional features at the Open Data Toolkit. P. 73-94 Available at:
https://idfi.ge/public/upload/Open%20Data%20Toolkit/Open_Data_Toolkit_ENG.pdf

- ◆ In case a public institution extracts data from a certain software, the information in the dataset should be organized accordingly. In particular:
 - First and foremost, the data should be imported in open formats (e.g. in Excel if data contains numbers).
 - Then, the dataset should be sorted in a way it contains the data the public institution intends to publish.
 - Public institution representative should make sure that the dataset is clear: there are no symbols, no numbers are omitted, no columns or rows are left out, the dataset does not contain any textual and stylistic errors.¹⁰

Upon processing a dataset, it is necessary to prepare data for its publication. At this point, one must once again check the accuracy of the data.



When **checking the quality of data**, attention should be paid to their completeness, cleanliness, and accuracy. Make sure that the information given in a separate dataset is based on the primary source, and the status of the document are specified (draft, revised or final version) (**data completeness**). Also, it should be checked that there are no blanks, incorrect numbers, symbols, or currencies in the dataset (**clean data**). The quality of the data is determined by their reliability. In addition, a dataset must enable aggregation and breakdown by particular criteria (**data accuracy**). At the same time, it is critical that the data prepared for publication are **updated**.

Technical Openness

At this stage, the data publication format should be determined. At least, the data must be published in Excel format, after which it is possible to use other formats (e.g.CSV).

Preferably, open data can become accessible through Linked Data and / or APIs. Linked data, APIs, and other open formats, as well as the technical features for data publication in these formats, are explained in Annex #3.

¹⁰See the data cleansing process in detail at the Open Data Toolkit. P. 73-94 Available at: https://idfi.ge/public/upload/Open%20Data%20Toolkit/Open_Data_Toolkit_ENG.pdf

Metadata

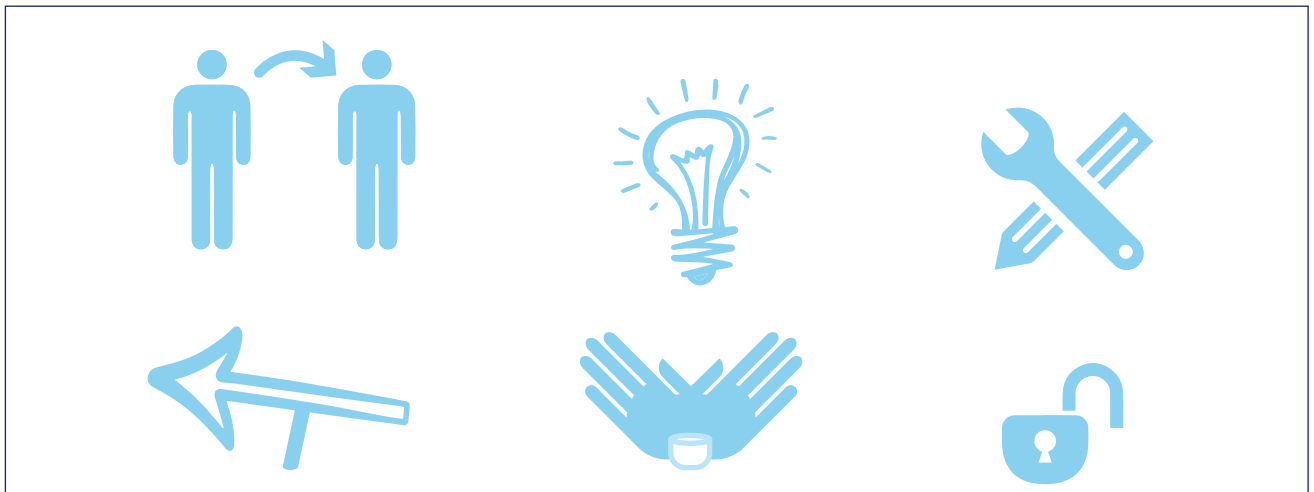
If a dataset contains complex and large volumes of information, it is crucial to add metadata (creation time, title, content / short description, author, type, size, update frequency, data collection method). Such details are important for any dataset, as users will easily get information about the dataset details and information displayed in it. Metadata should be simple and available in machine-readable formats. If it is published as linked data, the chances of discovering the data are significantly increased.

For more details about the standards and technical features of metadata, please check Annex #4.

Legal Openness

Open data publication standards should comply with the legal framework of a country. Datasets must be published under an open license so that third-parties could re-use and redistribute. In case of necessity, public institutions should have licensing procedures.

The Open Database License is a copyleft license agreement intended to allow users to freely share, modify, and use a dataset.



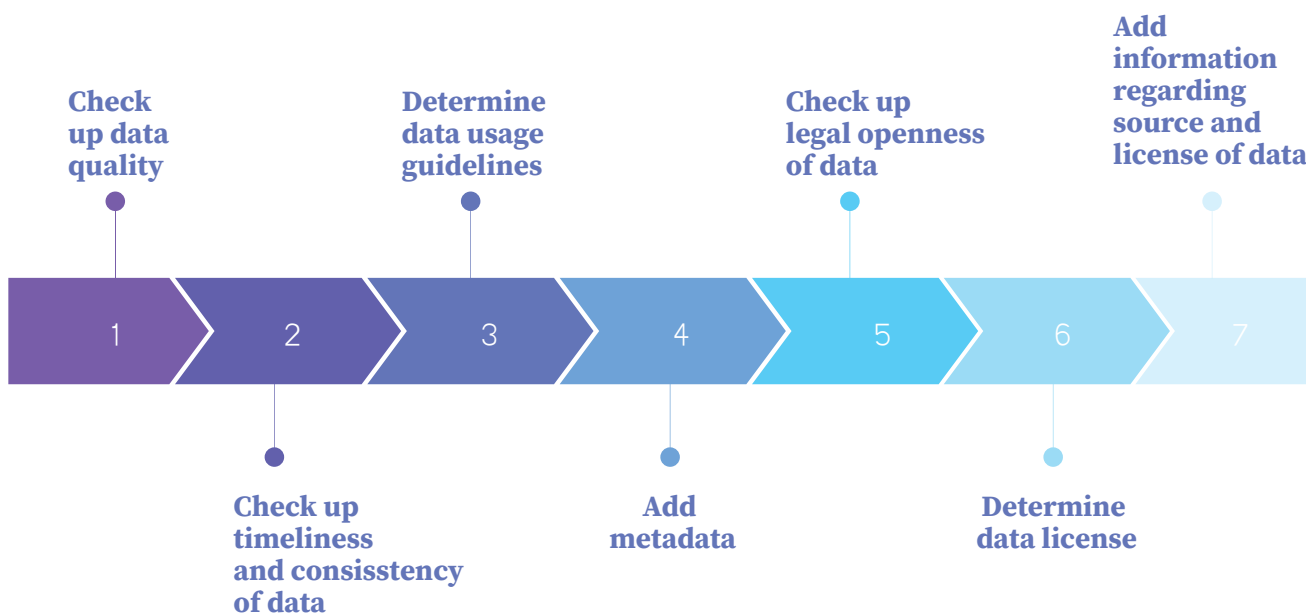
The dataset published on a website under an open license must be free for everyone. Also, an open license implies that regarding the data, anyone can:

- ◆ print out and share;
- ◆ publish in print or upload on another website;
- ◆ process, modify, and add (this does not imply changing or modifying the data);
- ◆ integrate with other data;
- ◆ display in other forms (text, video, graphics, etc.).¹¹

¹¹For more details, please visit: <https://opendefinition.org/guide/>

Public institutions often receive certain services from private companies. At this point, it is important for the signed contract include data collection and publication issues. In particular, the public institution should ensure that data is collected in open formats, as well as published/made available under the open license.

With all the above-mentioned issues in mind, the following steps must be taken prior to data publication:



3. Data Publication

Any public institution has various tools for publishing data, including a website, the central data portal, as well as the application program interface (API - Application Programming Interface). Publishing data on **a website of a public institution** is the easiest solution. In this case, public institution create a special section on their website, where datasets will be updated from time to time. This method is applied when a public institution does not have a large volume of datasets to be published. In technical terms, to this end, public institutions use Drupal, WordPress, or Django, the Content Management Systems (CMS).

Preferably, websites should have thematic sections, where the relevant type of information will be posted. Besides, it will be convenient for a user to search for specific information. For instance, a public institution may have one of the sections, **Budget**, where budget-related data will be published.

The publication of open data on **a unified data portal** is mainly chosen when public institutions are required to publish various datasets on different topics. Generally, in this case, a central government data portal is created and other public institutions also have the opportunity to publish important data on that portal.

In provide users with relevant tools for navigation on the portal easily, it is of great importance to create certain categories, among which the published datasets will be distributed. For example, the data published on the Central Data Platform of Georgia - data.gov.ge - are grouped according to public agencies and data formats, as well as the following topics: education and science, environment, geographical data, employment, economy, culture, governance and politics, population, construction, society, agriculture, social security, transport, tourism, finance, healthcare.

According to the best international practice, an online data portal is also created at the level of a particular city or municipality. On the example of the UK, the London Datastore¹² is a large open data portal with up to 1,500 open data themes. A similar practice can be seen on the example of the city of Toronto¹³. Through the online portal, active citizens, students, developers, business representatives, and other stakeholders have access to a variety of open data. The information on the portal is constantly updated and considers the needs of users. In this context, the case of Poland can also be mentioned - Otwarte Dane Gdynia¹⁴ - an open data portal that provides access to public data about the City of Gdynia.

When publishing data on an open data portal, it is recommended to indicate a small description, which will allow users to get familiar with the methodology used for data collection/processing. Also, a brief description may include additional information about the primary data as well as the data source.

As already mentioned, the Georgian Government created an open government data portal – Data.gov.ge and any public institution can make datasets available in open formats on this website, in addition to publishing them on its official website. To this end, public institutions can reach out to the **LEPL Digital Governance Agency**, which will provide consultation in publishing datasets on the platform. When identifying the datasets to be published on the portal, a public institution should give priority to the datasets particularly interesting for the private and civil sector representatives in terms of creating new services and platforms.

Publishing data through visualizations and graphics is sometimes important to provide information on certain issues to the public. It is known that a properly selected visual element, whether it is a graph, diagram, map, or other, makes the presented data easier to understand for a user. The content of the data must be taken into account when selecting the visualization. For example, while publishing numerical data (budget plan, execution, etc.), it is possible to use graphic diagrams, while visualizing location data – maps, etc.

¹² <https://data.london.gov.uk/topic>

¹³ <https://open.toronto.ca/>

¹⁴ <http://otwartedane.gdynia.pl/en/about/>

It should be noted that if a public institution decides to publish data in the form of visualization, it is also necessary to **publish the entire database** so that the user could do his/her analysis with the primary data.

Publishing data through an **Application Programming Interface (API)** tends to be a more technologically complex and convenient solution, as in this case, it is possible to link different systems and the data stored in them. Consequently, IT specialists have the unique possibility to automatically update data and directly use datasets of other systems.

Website	Data Portal	API
Easiest way to publish	Gathering numerous datasets in one place	Link and combine various Apps and systems
Drupal, Wordpress, Django	Classified datasets	Automatically update data

Sharing and promoting data - One of the main benefits of the data published in open formats is the possibility to reuse/redistribute them. Therefore, it is crucial to inform and encourage all stakeholders to use public data for new applications, services, and initiatives. To accomplish this, public institutions can cooperate with activists, bloggers, academia, non-governmental organizations, representatives of the private sector. At the same time, it is possible to organize data related events, such as hackathons, where data specialists, developers, activists, designers come together and create innovative products with the use of public data.

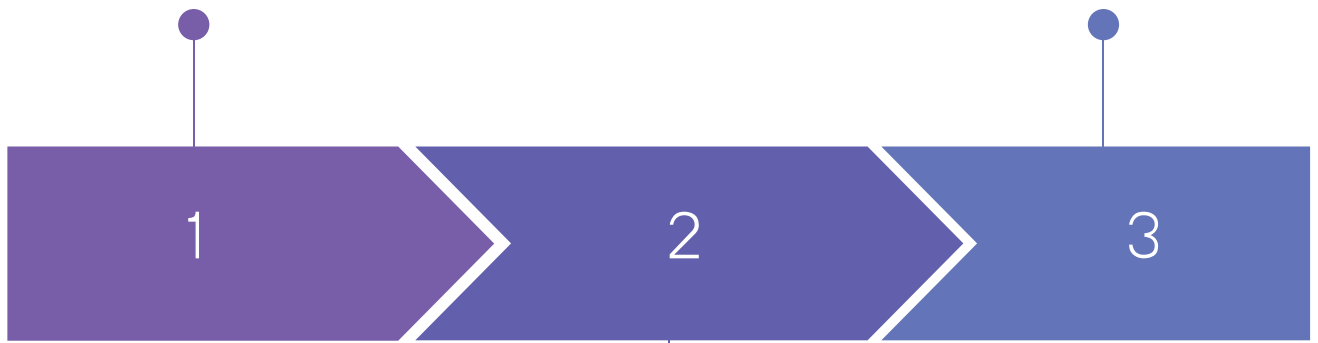
Together with data publication, it is no less important to store, maintain, and update data. First of all, representatives of a public institution should ensure that data, along with their metadata, are regularly updated. To this end, a public institution should have a predetermined frequency of data updates. In addition, each structural unit of the agency should have information about the data produced/stored by other structural units.

To avoid any technical problems on the website, **URLs** (URI and URL) of **datasets** should be regularly **checked** to make sure that the dataset address has not been changed or deleted. Also, if a public institution decides to update the website, make sure that previously published data addresses (URI / URL) are retained after the website update is completed.

To improve the quality of proactively available data, it is essential to take into account **feedback and comments** from users. A public institution should be open to consider users' views, improve a specific dataset, publish new data, or start collecting data of public interest for further publication.

Update data and metadata constantly

Improve data based on customer feedback



1

2

3

Check up URI and URL constantly



Practical Recommendations for Improving Open Data Management Practice

Practical Recommendations for Improving Open Data Management Practice

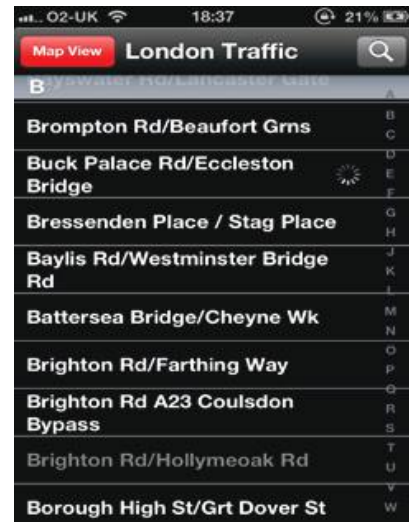
- ◆ A public institution should have a unified system/process of data management and publication, which will clearly define the role of particular structural units and the responsibilities of particular public servants.
- ◆ It is important to have a unified standard for producing/managing datasets. To this end, categories of data in a dataset and the frequency of its update must be determined in advance. The more detailed the data categories are, the easier it will be to compare, analyze, merge/generate new datasets, and use them for different purposes in the future (e.g., visualizations and development of new services).
- ◆ To increase access to open government data, it is important to frequently strengthen the capacity of public officials responsible for data management, and share new trends in this domain with them;
- ◆ When publishing data, the information must be accurate and up-to-date;
- ◆ When disclosing data, it is important to provide and/or publish it in the most open format available to a public institution. Unstructured formats should be used only in exceptional cases;
- ◆ The data should be published in the relevant thematic section of the website to ensure easy navigation on the website for a user;
- ◆ Datasets of economic importance and high public interest, together with the official websites of public institutions, should be additionally published on the central government data portal - data.gov.ge;
- ◆ To advance economic, social, and political benefits from open government data, the public sector needs to cooperate closely with a variety of stakeholders;
- ◆ A public institution itself should attempt to encourage the development of new services and platforms with the use of the data produced and/or published by it.

Annex #1: International Examples of Re-use of Open Data

This section of the guidebook provides examples of various interesting applications and services created by European countries developed through open government data. It is worth mentioning that **in all cases, the data through which a particular application or service was developed is constantly published/ updated by relevant public institutions and is available in open format.** Data publishing in an open format made it possible to create the following applications/services:

1. London Traffic – Great Britain

- ◆ The main function: it provides drivers with information regarding congestion on the roads, therefore assists to avoid traffic jams and save time;
- ◆ Data source: Cameras installed in the city and police data.



2. London Fire Brigade – Great Britain

- ◆ It is created by the British Fire Service. This application primarily shows the number of calls, and fire brigade arrival times. Data is broken down by city districts.
- ◆ Based on the data collected by the application, the fire department plans and implements various policies: decides where to deploy rescue brigades, how to equip a separate brigade, which is the highest risk area/place, in which part of the city it is required to conduct regular fire prevention inspections, to organize awareness-raising campaigns.



3. PulsePoint (Different countries, including the USA) and GoodSAM (London)

- ◆ A mobile app that helps people get first aid faster and, as a result, increases their chances of survival.
- ◆ In the USA, the app connects directly to 911 calls and indicates the location of the incident. In this way, a person in need may be assisted by someone around him who can provide first aid before an ambulance arrives.

PulsePoint

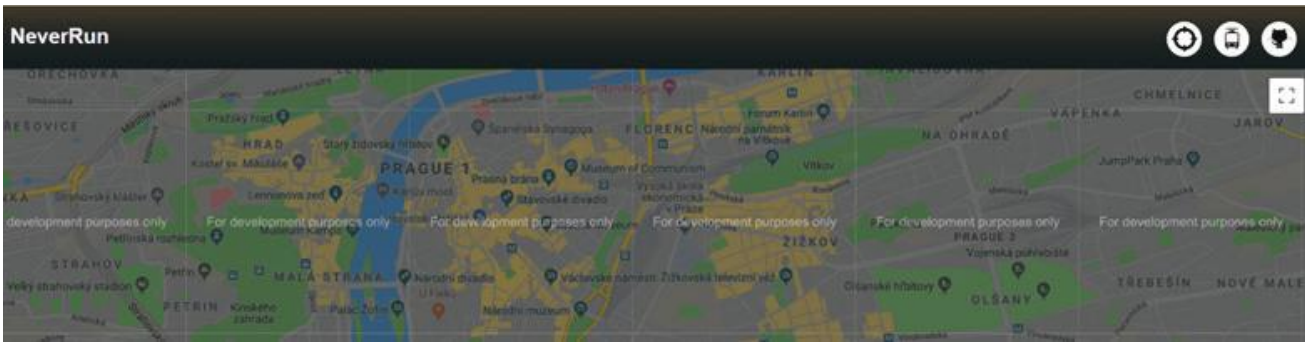
4. Ackee – LuckyMe! – The Czech Republic

- ◆ This is a Czech app that is available on iPhone and Apple Watch, and with its help, a user can prevent theft. In particular, the app evaluates the safety of city streets based on various data;
- ◆ The application shows the risk of crime in a particular district (on a scale of 1 to 5);
- ◆ If a user has detected different types of crime on any street (hooliganism, robbery, assault, etc.), he/she can also report the incident in the app;
- ◆ When evaluating streets, the app uses the following data: data from hospitals, municipal police, and consumers.



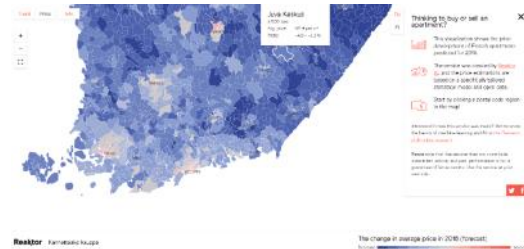
5. Never Run - Czech Republic

- ◆ The Czech app shows the location and schedule of Prague public transport;
- ◆ The application shows any type of transport around a user, who can choose the transport convenient for him/her to reach the destination;
- ◆ The application uses the data of the Prague City Hall, where the location of all types of transport is given in real-time.



6. Kannattaako Kauppa - Finland

- ◆ An open data-based interactive website in Finland offering apartment price rate forecasting service. Prices are given per square meter.



32

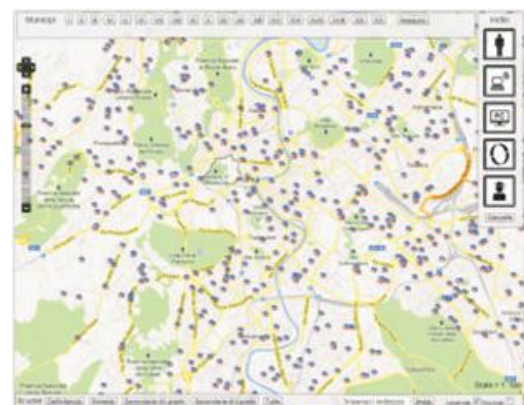
7. CitySDK Project - European Commission

- ◆ A project supported by the European Commission to assist cities to publish open data to allow developers to create mobile applications;
- ◆ As part of the project, to create applications, cities started to publish the following types of data: tourist destination addresses and entry time, historical site details, and transport schedule.



8. RomaScoula - Italy

- ◆ Through Macula App, educational institutions are compared in terms of teacher attendance, internet access, IT equipment, quality of education, diplomas awarded, and other indicators.
- ◆ Data source: Ministry of Education. Periodically updated according to the data received from the Ministry.



It is worth mentioning that open data use may also reduce the negative effect on the environment. For instance, Save the Rain¹⁵ is an app that shows users how to reduce their impact on the alarming decrease of annual rainfall forecast by the World Bank. The "rainwater harvesting" approach allows users to record the amount of rainwater that could be potentially saved.

The Danish electronic platform Huset's Web¹⁶ helps citizens to reduce their electricity consumption based on open datasets and architectural elements of Danish homes. Based on 200-year construction statistics and information generation on heating systems, a system for calculating the optimal electricity consumption has been created, which reduces utility bills for each citizen using the system (the citizen can add information about the building as well as the heating system).

The benefits of open data products can be assessed in two ways. The first is when developers use open data to create a new and valuable product/service, and the second is when they combine open data with an existing product or service. In the ideal scenario, we get a "cycle" where a new product increases the demand for open data, and vice versa, the data indicate the need to create a new product/service. For example, Dataviva¹⁷ is the largest open data portal in Brazil. The platform includes government data for the last decade of the country by various categories, such as economy, education, industry, labor market, and others, with the indication of locations. Providing around billion visualizations, the electronic platform presents new needs and facilitates dialogue between the public and private sectors.

Another relevant example is the dataset available on **London Open Portal**¹⁸ containing information about bicycle rental (indicating location) and the number of available bicycles (excluding sealed and/or faulty bicycles). Moreover, the same platform reflects the intensity of bicycle rental by the population throughout the year. Analyzing this data will help the entrepreneur choose the right location/district when opening a bike shop.

The case of the municipality of Nijmegen in the Kingdom of the Netherlands is a great example to illustrate how to create a useful service for citizens using open government data. The municipality uses its open data portal¹⁹ to create a map that is accessible to every citizen and shows train stations, bus stops, parking lots, etc. The municipality also updates the information according to the season, for example, during the winter season, a map shows which road has been and/or is planned to be salted.²⁰

¹⁵ <https://www.savetherain.org/>

¹⁶ <https://www.modstroem.dk/energiberegner/>

¹⁷ <http://dataviva.info/en/about/>

¹⁸ Cycle Hire Availability. London DATASTORE. Available at <https://data.london.gov.uk/dataset/cycle-hire-availability>

¹⁹ <https://opendata.nijmegen.nl/>

²⁰ *The Economic Impact of Open Data Opportunities for value creation in Europe. (European Data Portal, 2020. p. 127) Available at <https://www.europeandataportal.eu/sites/default/files/the-economic-impact-of-open-data.pdf>*



Annex #2

Annex #2: Legislative Framework and Practice of Access to Public Data in Georgia

Access to Public Information

According to Article 24 of the Constitution of Georgia, every person has the right to freely receive and disseminate information, to express and disseminate his opinion orally, in writing, or by other means. Article 41 of the Constitution of Georgia stipulates that every citizen of Georgia has the right to get acquainted with the information contained in state institutions, as well as the official documents if they do not contain state, professional or commercial secrets. In addition, Article 37 of the Constitution of Georgia guarantees the human right to receive complete, objective, and timely information on the state of his or her work and living environment.

Georgia does not have yet an independent legal act on Freedom of Information. In 2014, the Government of Georgia, within the framework of the Open Government Partnership (OGP) Action Plan, undertook an obligation to develop a new law regulating freedom of information²¹. Later, this commitment was also written in the Association Agreement between Georgia and the European Union and the 2016 National Action Plan²² for the implementation of its agenda. The bill on freedom of information is planned to be initiated in the Parliament soon.

Issues regulating access to public information are given in the General Administrative Code of Georgia. **The General Administrative Code** clarifies that public information is: information held in a public institution, as well as the information received, processed, created or sent by a public institution or employee in connection with official activities, including drawing, model, plan, diagram, photograph, electronic information, video and audio recordings, and also information proactively published by a public institution.²³

According to the definition of public information, the **obligation to provide information** extends to public institutions, such as administrative bodies and private legal entities financed from state or local budget funds.

It should be noted that the legislation of Georgia includes the presumption of publicity of information protected in a public institution. Any information stored in a public institution constitutes public information unless proven otherwise. Public institutions have the right to refuse to disclose information only if they substantiate that the information contains personal data, state, or commercial secrets.

²¹ Approved by the Resolution №557 of the Government of Georgia of September 18, 2014, see Commitment 10, Available at: <https://matsne.gov.ge/ka/document/view/2510377>

²² Approved by the Decree of Government of Georgia 7382 of March 7, 2016, see Article 70.10, Available: <https://matsne.gov.ge/ka/document/view/3222307>

²³ General Administrative Code of Georgia, Article 2, paragraph 1, sub-paragraph "m".

According to the current legislation, a public institution is **obliged to provide the requested public information immediately**. In case the requested information is large, needs to be processed or stored in another administrative body, public institutions are given 10 working days to provide the information²⁴.

The release of public information is free. Public institutions have the right to request a fee for copying public information or a fee for transferring it to a compact disc.

The issue of personal data protection in the process of processing and publishing open data - According to the Law of Georgia on Personal Data Protection, personal data is any information related to an identified or identifiable private person. A person is identifiable when it is possible to identify him or her directly or indirectly, in particular by an identification number or by a person's characteristic physical, physiological, psychological, economic, cultural, or social characteristics. For example, personal data is a person's name, surname, personal number, photo, video recording, email address, bank account number, personal correspondence, etc.²⁵

This law sets out the rules, principles, and security norms that apply to the processing of personal data in any form.

According to the General Administrative Code of Georgia, a public institution is obliged not to disclose the personal data of a person without the consent of that person, except when required by law to ensure state or public safety, protection of public interests, health, or the rights of others.

It is noteworthy that personal data of public officials (as well as the candidates) are public. The definition of public officials together with a detailed list are outlined in the Law of Georgia on Conflict of Interest and Corruption.

To fully neutralize the risks and security norms in the field of personal data processing, as well as taking into account good practice, the institution may designate a person responsible for the protection of personal data policy.

The role of the **public information register** in protecting the principle of publicity of information is important. According to the current legislation, a public institution is obliged to enter the public information in that institution in the public register.²⁶ This refers to maintaining a public information register by public institutions, in which the public information stored in the institution is registered with an indication of the name, receipt,

²⁴General Administrative Code of Georgia, Article 40, Paragraph 1.

²⁵Office of the State Inspector. "What is my personal data?" Available: <https://personaldata.ge/ka/ra-aris-personaluri-monacemebi>

²⁶General Administrative Code of Georgia, Article 35.

creation, processing, or publication of public information. The current legislation defines the authority of administrative bodies to use software and unified automated means of access to document management and access to information, including storing and issuing in electronic form any document created or maintained by them.²⁷

²⁷General Administrative Code of Georgia, Article 351



Adopting Unified Standards for Access to Open Data

Adopting Unified Standards for Access to Open Data

The current version of the **Bill on Freedom of Information** defines the concept of open data and clarifies that it is freely accessible data that can be processed and reused for any purpose and without prior consent. However, after the adoption of the bill in the given form, the administrative bodies will be obliged to publish the open data available to them in a machine-readable format on the open data portal and to further update them.

In 2015, within the framework of the Open Government Partnership Action Plan approved by the Government of Georgia, the Legal Entity of Public Law (LEPL) of the Ministry of Justice, the Data Exchange Agency (now the Digital Governance Agency) established an open government data portal - www.data.gov.ge, where public institutions have the opportunity to publish the data they produce in open formats.



Rule on Proactive Disclosure of Public Information

Rule on Proactive Disclosure of Public Information

The current legislation imposes an **obligation on public institutions to proactively disclose information** and outlines that proactive disclosure implies the publishing of public information of public interest on the electronic platform by public institutions.²⁸ Nevertheless, the law does not specify the list of information that public institutions are obliged to publish. According to the General Administrative Code, the list of information to be proactively published is determined by the by-laws of the administrative bodies. Accordingly, the determination of the standard of information to be proactively published is at the discretion of the administrative bodies. In practice, administrative bodies have set different standards for the information they proactively publish. This is especially noticeable in the cases of local self-government bodies.²⁹

The current legislation defines a unified list of information to be proactively published for the Government of Georgia, ministries, and their subordinate public institutions (LEPLs and sub-institutions). The list includes the minimum mandatory information that ensures proactive transparency of the activities, plans, financing, and cost of administrative bodies, public procurement, and property privatization, as well as other important information.³⁰

²⁸ General Administrative Code of Georgia, Article 28, Paragraph 2.

²⁹ Local Self-Government Index, Unified National Assessment of Georgian Municipalities, p.11, available at: <https://bit.ly/2IEYR5E>

³⁰ Resolution of the Government of Georgia № 219 On Requesting Public Information in Electronic Form and Publishing It Proactively. August 26, 2013 <https://matsne.gov.ge/document/view/2001875?publication=0>



Unified State Registry of Information

Unified State Registry of Information

The Law on the Unified State Registry of Information³¹ has been in force in Georgia since June 2011. Accordingly, the Unified State Registry of Information is a unified catalog of databases, registers, services, and information systems of the subject of the registry, which aims to describe information resources in the Georgian public sector, set common standards for working with information, coordinate information systems and promote consistent development, promotion, and efficient use of public resources.

The law sets out the obligation to register significant changes, extensions, combinations, cancellations, destruction, archiving, and transfers of the registry, database, service, or information system. In particular, according to the law, a public institution is obliged to notify the LEPL Digital Governance Agency in writing no later than 30 days after the creation of the database or register, and 30 days before the cancellation.

The written notice for the creation of a database, registry, services, and information system to the Digital Governance Agency shall contain information describing the operational and information-technological circumstances. In particular, the information to be provided about the operating issues should include the following data:

- A) Contact information of the data owner and / or his / her authorized person and description of its main activity.
- B) A description of the database, register, service, or information system and the purpose of its production.
- C) The procedures used to produce a database, register, service, or information system, a brief description of the information processing business process, and other information describing the production process of the database, register, service, or information system.

Information to be provided on the **information-technological circumstances** should include the following data:

- A) Technical standards applicable to databases, registries, services, or information systems;
- B) Contact details of the database, registry, service or information system administrator or another authorized person;

³¹ Rule of Georgia regarding On the Unified State Registry of Information: <https://www.matsne.gov.ge/ka/document/view/1338521?publication=1>

C) The data of the person/body or authorized person responsible for identifying and correcting the errors made in the production of the database, register, service, or information system if any;

D) Data protection measures.

The Digital Governance Agency also makes recommendations to registry subjects on the creation, use, modification, extension, integration, compatibility, deletion, and other issues regarding a database, registry, service, or information system.

According to the same law, the data stored (processed) in the database, register, information system, and in the process of providing services of the subject of the register are public and available, unless the access to or publication of the data is restricted by law.



Annex #3

Data Formats

Annex #3: Data Formats

Types of formats

Data can be available in a variety of formats. Here are some of the most commonly used formats, their features, and functions:

JSON

A simple file format that you can read in any programming language. Simplicity means that it is easier to work on a computer than with other formats, such as XML.

XML

XML is a widely used format for data exchange because it gives good opportunities to keep the structure in the data and the way files are built on and allows developers to write parts of the documentation in with the data without interfering with the reading of them.

RDF

An internationally recommended open format. RDF makes it possible to represent data in a form that makes it easier to combine data from multiple sources. RDF data can be stored in XML and JSON, among other serializations. RDF encourages the use of URLs as identifiers, which provides a convenient way to directly interconnect existing open data initiatives on the Web. RDF is still not widespread, but it has been a trend among Open Government initiatives. The inventor of the Web, Tim Berners-Lee, has recently proposed a five-star scheme that includes linked RDF data is one of the front runners in terms of openness.

Spreadsheets

Many public officials have information stored in spreadsheets, such as Microsoft Excel spreadsheets. After correctly describing the values of each column, this data can be used immediately.

However, in such files, you can often come across formulas that are difficult to work with. Therefore, it is best to keep descriptions of such calculations with spreadsheets as it will be easier for users to read them.

CSV (comma-separated values)

CSV is a very useful format because of its compact size, which allows you to convert a large data set retaining the same structure to another format. To read the data, you need the accompanying documentation, which explains what kind of data is presented. Furthermore,

the structure of the file must be respected, as a single omission of a field may disturb the reading of all remaining data in the file without any real opportunity to rectify it because it cannot be determined how the remaining data should be interpreted.

Text Document

Classic documents such as Word, ODF, OOXML, or PDF can be enough to display multiple types of data. Their distribution will be cheap because data is often produced in this format. The format does not provide control over the structure, making it often difficult to enter data automatically. Be sure to use templates as the basis of documents.

Besides, the maximum use of identical typography markup will be useful for further use of the data as it makes it easier for the electronic device to separate the headings from the information. It is usually not recommended to upload data in a text editor format if it is stored/produced in another format.

Plain Text

Computers can easily read documents in plain text (.txt) format. However, they generally exclude structural metadata from inside the document, therefore, developers will need to write a parser to read a document.

Scanned image

It is the least suitable form for most of the data. However, TIFF and JPEG, at a minimum, allow you to at least mark them with documentation of what is in the picture. This can be useful for displaying data as images whose data were not created electronically (an obvious example is old records and other archival material) - a picture is better than nothing.

Proprietary formats

Some dedicated systems have their data formats that they can save or export data in. It can sometimes be enough to expose data in such a format - especially if it is expected that further use would be in a similar system as that which they come from. Where further information on these proprietary formats can be found should always be indicated, for example by providing a link to the supplier's website. Generally, it is recommended to display data in non-proprietary formats where feasible.

HTML

Nowadays data is often available in HTML format on various websites. For infrequently variable and small volumes of data, it will suffice to place them in this format. Also, publishing data in this format is cheap and easy to find/share because it is uploaded on a

certain page on the website. However, downloading data in this format directly and linking it to other data is inconvenient.

Typically, it would be most appropriate to use tables in HTML documents to hold data, and then the various data fields must be displayed and are given IDs which make it easy to find and manipulate data.

Open file formats

Even if the information is provided in an electronic, machine-readable format, and in detail, problems may arise due to the format of the document itself.

The formats in which information is published - in other words, the digital base in which the information is stored - can be either "open" or "closed". An open format is one where the specifications for the software are available to anyone, free of charge so that anyone can use these specifications in their software without any limitations on re-use imposed by intellectual property rights.

If a file format is "closed", this may be either because the file format is proprietary and the specification is not publicly available, or because the file format is proprietary and even though the specification has been made public, re-use is limited. If information is released in a closed file format, this can cause significant obstacles to reusing the information encoded in it, forcing those who wish to use the information to buy the necessary software.

The advantage of open format data is that developers can create several automated programs and services that retrieve data from documents published in open formats. This solves all the problems of using information from these files.

Using closed file formats whose specifications / technical requirements are not available, may result in dependence on third-party software or format licensees. In the worst case, this could mean that the information can be read by only one program, which can be expensive, outdated, or only available on one platform.



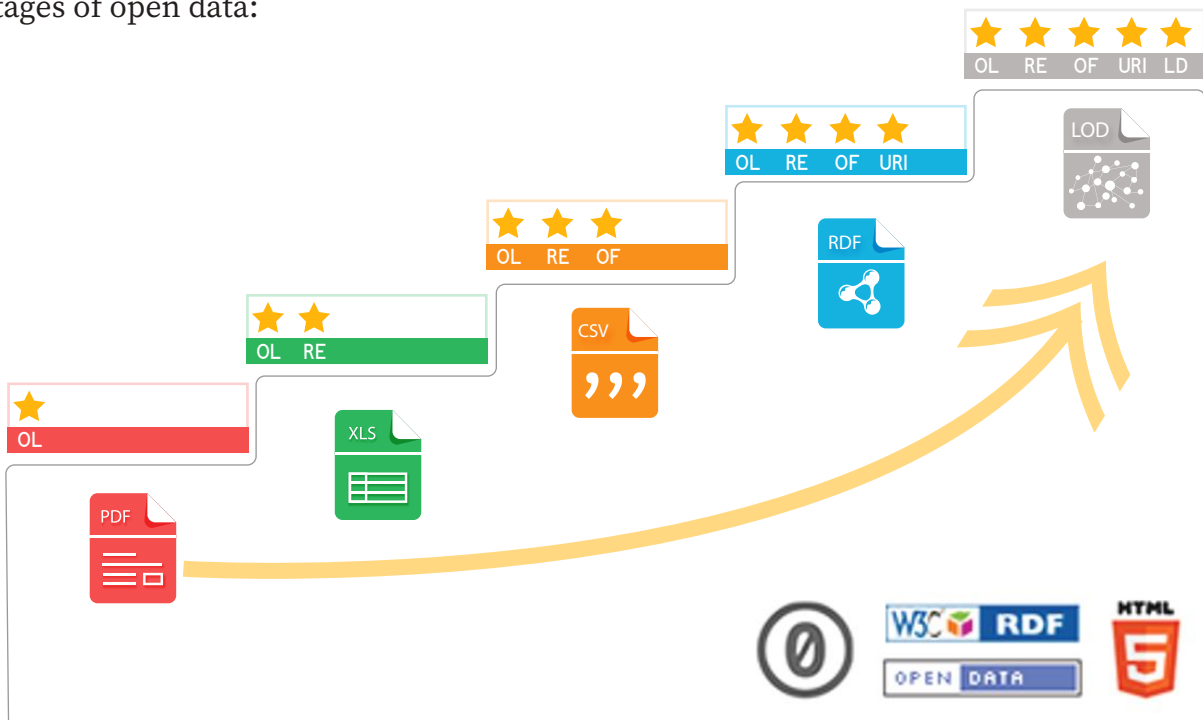
The 5-star open data model



The 5-star open data model

The W3C (World Wide Web Consortium) has created an open database model based on the degree of openness: a 5-star open data model.

5 stages of open data:



★	<p>Publish data in any format online, under an open license. Example: https://5stardata.info/en/examples/gtd-1.pdf</p>
★★	<p>Publish data online in a structured way (editable by both electronic and human devices, for example, Excel, spreadsheets, and not spreadsheet photos). Example: https://5stardata.info/en/examples/gtd-2.xls</p>
★★★	<p>Publish data in open formats. Same as the 2nd, but already in non-commercial, non-proprietary formats (e.g. *.CSV instead of *.XLSX, *.ODT instead of *.DOCX, ODP instead of PPTX, etc.). Example: https://5stardata.info/en/examples/gtd-3.csv</p>
★★★★★	<p>Same as 3rd, but already using W3C standards (RDF and SPARQL) URI is used to reference the data to make the data searchable. This is the first step towards Linked Data. This means you can convert data to RDF format and enrich your metadata with a URI. Example: https://5stardata.info/en/examples/gtd-4/</p>



Same as the 4th, but already with the ability to add contextual data that is indicated/given in the data employing identifiers. Accordingly, the published data are related to other relevant data.

Example: <https://5stardata.info/en/examples/gtd-5/>

The main strength of open data is that it is easy to re-use and facilitates the interoperability of systems and services. In general, data formats fall into two categories:

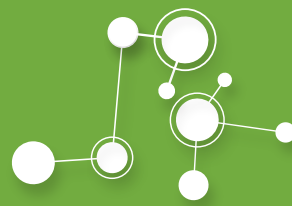
1. Structured data (readable by both humans and electronic devices);
2. Unstructured data (readable only by humans).

JSON	Shapefile	RTF (for texts)
XML	GeoJSON	HTML
RDF	GML	Excel
CSV	KML	PDF (for texts)
TSV	WKT	
ODT	KMZ	

A shortlist of electronically processable formats

It is recommended to convert your existing data into open and machine-readable formats to obtain high-quality Linked Data.

Electronically Processable Format	Electronically Processable Geodata	Less Readable	Closed
JSON	Shapefile	PDF (for texts)	Images (PNG, JPG)
XML	GeoJSON	HTML	Charts
RDF	GML	Excel	
CSV	KML	Word	
TSV	WKT		
ODT			
Technical openness of data			



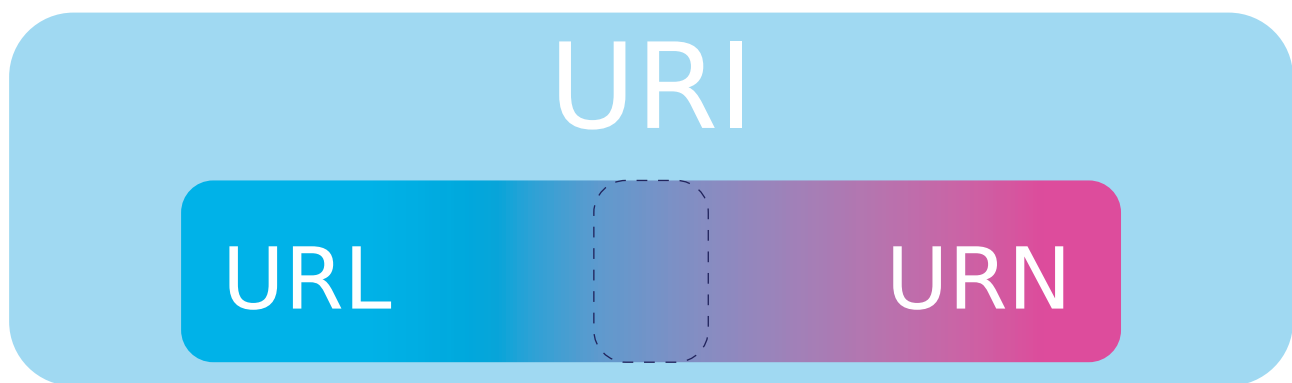
Linked Data



Linked Data

The concept of Linked Data enhances the interoperability and discoverability of datasets. Linked Data is not the same as open data. While open data refers to the openness of data itself, Linked Data is a way of publishing Open Data as Linked Data and enriching datasets with Linked metadata. This is where the data becomes more technical/dynamic.

Linked Data are pieces of information that are linked through a graph connection. Opposed to other relational descriptions of data, in Linked Data, a machine can walk through the graph and understand the content. This is considered a revolution in the area of data storage and sharing: the computer can, to some extent, qualitatively interpret data. This is possible because the data is enriched with uniform descriptors. By means of these descriptors, the data is no longer static content, but is described, and can therefore be interpreted, regardless of any distinguishing factor (language, file type, etc.).



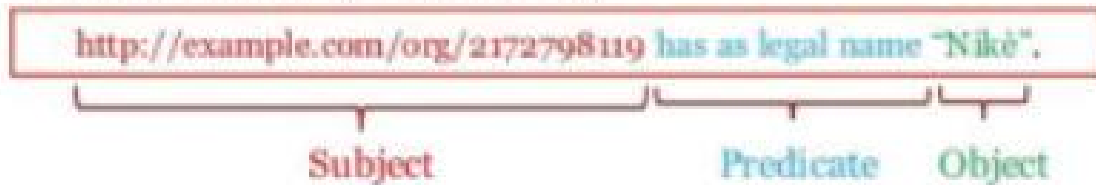
When publishing Linked Data, please take into account the following aspects:

RDF (Resource Description Framework) is the basic principle of Linked Data. This is the new general structure/syntax for representing data on the web. This syntax is a link (URI - a unique identifier for a resource) that is made up of 3 descriptors, which altogether are called a Triple. By describing an object with this triple, it becomes linked. Whereas terms may differ from sentiment, the structured way of describing them through the RDF triple overcomes this. Also, as many terms are described through RDF terms, they can be linked to each other.

Every piece of information expressed in RDF is represented as a **triple**:

- **Subject** – a resource, which may be identified with a URL.
- **Predicate** – a URI-identified reused specification of the relationship.
- **Object** – a resource or literal to which the subject is related.

Example: name of a legal entity:



The photo explains the idea of the RDF Triple

RDFa

The frequently used RDF technique on data portals is RDFa: embedding RDF in HTML.³² It is recommended to publish metadata with RDFa embedded in HTML.

```
<html>
<head> ... </head>
<body>
...
<div resource="http://example.com/org/2172798119"
typeof="http://www.w3.org/TR/vocab-regorg/RegisteredOrganization">
<p>
<span property="http://www.w3.org/TR/vocab-regorg/legalName">Nike<span>
Address: <span property="http://www.w3.org/ns/locn#fullAddress"> Dahlistraat
24, 2160 Wommelgem </span>
</p></div>
</body>
```

embedding RDF data in HTML

URI

URI is a unique resource identifier and can refer to text, Unified Resource Name (URN), or Unified Resource Locator (URL). A URI is a sequence of characters that identifies an abstract or physical resource. Its main function is to identify something. Typically, in the case of Linked Data, URIs are triples in the form of a URL (`http://www.europeandataportal.eu/`) or vocabulary specific identifiers.

³²For more information on RDFa see: https://www.slideshare.net/fabien_gandon/rdfa-in-a-nutshell-v1

A URL is a URI that, in addition to identifying a resource, also provides information about the location of that resource. A URN is a URI that only identifies a resource in a specific namespace (and therefore in a certain context), but does not indicate its location. For example, URN `urn:ISBN:0-395-36341-1` is a URI that points to a resource (book) 0-395-36341-1 in the ISBN namespace, but, unlike the URL, the URN does not indicate the location of this resource: it does not say in which store you can buy it or from which website you can download it.



Annex #4

Metadata Standards

Annex #4: Metadata Standards

There exist metadata standards that provide an overall structure and contribute to ensuring system compatibility and improving the discoverability of data. You can see the example of metadata based on the dataset from the London Data Portal:³³

Title of Metadata	Bicycle Rental Statistics - Santander Cycle Hire Scheme
Short title of metadata	Bicycle Rental Statistics
	Transport
Sub-category	Cycling
Description	A total number of bike rentals by the day, month, and year according to the Santander Cycle Hire Scheme
Publisher	GLA
Accessible Data	31 July 2010 - 31 August 2020
Author	London Transport (TFL)
Creation Date	20 September
Coverage Area	London
Coverage Time	Daily, weekly, annually
Type	Administrative

³³The database is available at: <https://data.london.gov.uk/dataset/number-bicycle-hires>

Language	English
Coverage Time	London Transport (TFL)
Warning/Note	Please note that considering the information provided by the system as of August 1, 2017, there is a small error in the daily hiring rates.

Here are the most popular metadata standards used in the Open Data context:

- ◆ Common European Research Information Format (CERIF) is an EU recommendation to the Member States and is used for all research information;
- ◆ Dublin Core (DC) - used to describe webpages;
- ◆ Data Category Vocabulary (DCAT) - used for datasets on the internet, also based on DC;
- ◆ Comprehensive Knowledge Archive Network (CKAN) - governments in different countries use for open data platforms;
- ◆ eGov metadata Standard (eGMS) - based on DC;
- ◆ Inspire - used for datasets containing geospatial coordinates.

The Best Practice of Metadata

Providing high-quality metadata is a big challenge but a necessary practice. According to W3C guidelines and best practices, the use of the DCAT-AP for providing metadata is recommended. Also, it is significant to publish the metadata with the data using a machine-readable format and standard terms to define the metadata.



Further Reading

Further Reading

1. Institute for Development of Freedom of Information, Open Data Handbook Available at: <https://datalab.ge/ge/toolkittext/toolkit/5/>
2. Transparency International Georgia, Determination and Measurement of Open Data. 2014. Available at: <https://bit.ly/3kUc0Sq>
3. Open Knowledge Foundation, Open Data Handbook. Available at: <http://opendatahandbook.org/>
4. Eight Principles of Open Government Data. Available at: <https://opengovdata.org/>
5. European Data Portal. How to build an Open Data Strategy. Available at: <https://www.europeandataportal.eu/en/providing-data/goldbook/how-build-open-data-strategy>
6. Capgemini Consulting. Creating Value through Open Data. 2015. Available at: <https://bit.ly/2rRoygs>
7. Open Knowledge Foundation. The Open Data Handbook. Available at: <https://okfn.org/opendata/>
8. Data Opening. Good Practice Guide. Available at: <https://bit.ly/3kWLQhZ>

