Data opening

Good Practice Guide

Fundusze Europejskie Polska Cyfrowa

Ministerstwo Cyfryzacji

OTWARTE DANE

Unia Europejska
Europejski Fundusz Rozwoju Regionalnego

Table of contents

If you are considering data opening in your government office, or if you have already started this process and are looking for ways to improve it (chapter 6), be sure to read on. I**n this guide, we share our experience in public data opening**, as well as present insights from local authorities with whom we conferred.

Data opening is a process without a clear beginning or end, and one in which much depends on factors such as legal regulations, the institutional environment, the level of cooperation between public institutions, and the general attitude of civil servants towards transparency of public life. This guide describes the basic framework for this process, referencing relevant legal acts, identifying desired institutional settings, and presenting practical scenarios for data opening in government offices.

This guide was written for managers and employees in central and local government administration offices, representatives of non-governmental organisations and all who base their knowledge of public administration activities on data generated and processed in public institutions. It is focused on covering the most important legal regulations affecting the opening and use of public data.
It shows inter-institutional and non-institutional cooperation models which have proved to work best in this context. You will also learn how to implement the data opening process effectively. If you are looking for technical guidance, you will find it in the standards for data openness: **technical**, **API [Application Programming Interfaces]** and **security**.

1. What is data and why it is worth using it?

We shall start by defining the word **data**. According to the Merriam-Webster dictionary, **the definition of data** is

Dane to:

1. „factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation";
2. „information in digital form that can be transmitted or processed".

In the governmental **Public Data Opening Programme** (hereinafter: Programme) you will find a definition of public data, which is data possessed by public offices, regardless of who generated such data.

Public data - numbers and singular events or objects at the lowest possible level of aggregation, which have not been processed by public administrations into reports, charts, etc. and have not been given an appropriate context or interpretation.

It is difficult to enumerate all data used by an administration in their daily tasks. This data include information on income, expenses, education and the education system (including exam results), level of air pollution, local business registers, libraries, tourist attractions, building permits, hotels, water systems, airports, and many more.

Recommendations

Analyse the most popular resources used by external entities.

Citizens have a right to know about each public action, and using data can positively affect their quality of life.

For example, in Spain there has been much interest in data on petrol stations and the prices offered by them, and in Poland, sites processing communication data are thriving. Sites using the data published online, such as those published in the National Court Register, are also popular. This perfectly illustrates organisations' and companies' ability to build useful products from publicly available data.

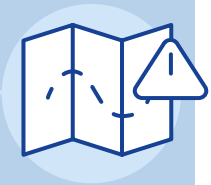
Public data has many uses, and its opening enables the following goals to be achieved:

Improving public institutions' work and facilitating the work of civil servants

As described in the Programme: "the public administration does not always see itself as the main beneficiary of data opening. The opening of public data improves the exchange of knowledge and information between government units and communication with citizens. The data can be used by other public entities to create innovative services for citizens". As shown by the Ministry of Finance, opening data can even play a role in the fight against tax crimes. There are many examples of using data in administration, ranging from complex databases, such as PESEL or CEPiK, to less complex ones, e.g. the registration of requests to the office for access to public information.

Data opening is also beneficial to authors of policies, strategic documents and drafts of legal acts. However, the lack of reference to specific data, and analysis based on such, remains a great drawback of Impact Assessments. Access to data can change this situation.

Increasing citizens' involvement in public affairs



Sharing data can improve knowledge about the activities of legislative and executive authorities, and thus provide more transparent access to law. **Data downloaded from the Parliament** allows you to follow public debate and **provides information about your representatives**. It is difficult not to agree with the authors of the **National map of security threats**: "a sense of security is one of the basic needs of every human being and is a crucial influence on the functioning of every community. For this reason it is important to create tools which will allow reliable and clear identification and presentation - including presentation to local communities - of the scale and type of threats, as well as the institutions which share the responsibility of ensuring public order and safety. Data also builds knowledge about local and national problems and phenomena, e.g. ecological or social."

Building and innovative environment

Open data supports the growth of business and start-ups, as well as the development of the labour market. According to data from **the European Commission**, the market for services and products using open data is expected to grow to EUR 75.7 billion in 2020. The number of people working in sectors based on open data is expected to increase to around 100,000 by 2020. For the time being, accelerated development of technological innovations can be observed abroad, where services such as websites based on the **labour market data** or supporting **fire department operations** exist. In Poland, websites presenting timetables or information on business partners remain the most popular.

Solving social and economic problems

It is necessary to look upon the process of data opening holistically, as part of the implementation of State strategy. The issues of data opening and facilitating their re-use were included in, e.g. **Strategy of Efficient State 2020 and Strategy of Innovation and Economic Efficiency "Dynamic Poland 2020"**.

Based on public data, problem-solving strategies are developed at local and national level. In addition to the safety and health issues mentioned above, you can distinguish those categories of data that allow you to drive rational **education policy, sports promotion** or **economic development**. Technological solutions alone will not cause problems typical for a given area to disappear. According to **OECD**, (...) "the challenge is not to introduce digital technologies into public administrations; it is to integrate their use into public sector modernisation efforts."

And the wide use of public data, especially open data, by the administration is one of the conditions for improving its effectiveness. Why is it so important? In order for the actions of the State to be effective, these actions must be confronted at every stage, from planning through monitoring to evaluation, with facts , i.e. data about the intensity of a given phenomenon and the changes that have taken place. What is more, open access to such information increases the transparency of administrative activities and enables the involvement of scientific, non-governmental and expert communities to cooperate in solving socio-economic problems and shape public policies.

Do you check timetables in the application; do you use navigation? These are the results of open data. Open data means greater transparency of State actions and better quality of public policies. Open data helps to develop business and facilitate the work of scientists.

2. What does open data mean?

Data, in order to effectively serve the purpose of achieving the above objectives, should:

1. be prepared in a user-friendly way - the key is a communicative language that is understandable for the recipient;
2. fulfil all the pillars of openness, in which reference is made to legal matters, shared in a machine-readable format,[1] in open formats[2] that can be re-used.[3]

More detailed features of openness can be found in data openness standards: **technical**, **API** and **safety** in the Programme that refers to **international standards**, including those developed by the **European Commission**, which specifies that data are:

1. **available** - accessible without any limits and for any purpose to a broad range of users (in particular to citizens, companies, universities, institutions);
2. **publicised in the source version** - available in the original, unchanged version, and not as analyses, summaries, or abbreviations, so that it is possible, for example, to combine data from different sources;
3. **complete** - the full version should be shared;
4. **up-to-date** - the data should be shared at a time when they are still valuable;
5. **machine-readable** - available both in machine-readable formats and open format. Examples of such formats are CSV, XML, spreadsheet. Formats that are typically difficult to machine-read

---

[1] **Machine-readable format** - file format organised in a way that allows computer applications to identify, recognise, and retrieve specific data, including individual statements of facts and their internal structure.

[2] **Open format** - file format that is not associated with the platform and is made available to citizens without any restrictions that would hinder the re-use of documents.

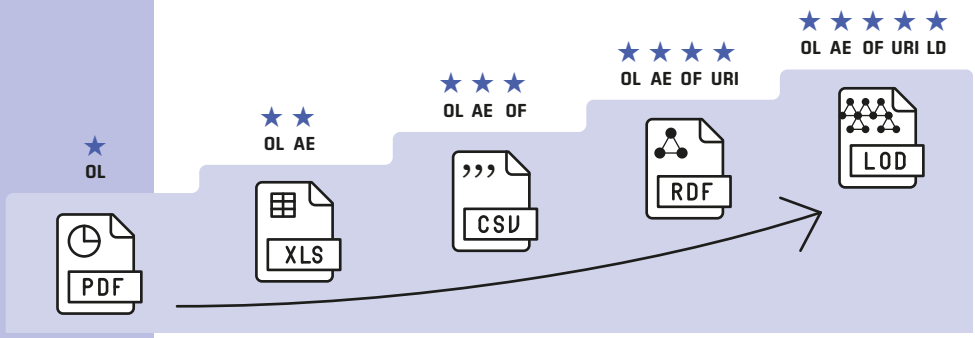
[3] Later in this guide you will find more information about reusing data, including definitions and rules.

are PDF, HTML, or text files, which become suitable for re-using only after they have been converted into one of the open formats;

6. **shared indiscriminately** - available to anyone without registration, verification of identity through signing in with a password, login, and without having to sign any contracts;
7. **available without any legal restrictions** - data cannot be subject to copyrights, patents, trademarks, or trade secrets and may be used for any purpose without the need to obtain permission to use it;
8. **not-restricted** - available in a widely used format which is not controlled by any entity.

To capture the difference between open data and data to which access - and therefore the possibility of using them - is difficult, please refer to the **5 levels of openness system**.

The system was designed by the Internet precursor, Tim Berners-Lee. As shown below, merely placing a file with information on the web fulfils only the first of five levels. It is worthwhile to enable residents, business representatives, students, or non-governmental organisations to use further, higher levels of openness, where the particular asterisks mean:



Źródło: http://5stardata.info/en/

where the particular asterisks mean:

- ★ Sharing data on the Web (in any format) under the terms of an open licence
- ★★ Sharing data in a structured form (e.g. a spreadsheet instead of a scanned image of a table)
- ★★★ Using open formats (e.g. CSV instead of a spreadsheet)
- ★★★★ Using URI to mark resources to make them searchable
- ★★★★★ Connecting data to provide context

★ Sharing data in the form of a table in a PDF file

TABL. 4. **NAJDŁUŻSZE RZEKI**
THE LONGEST RIVERS

Rzeki Rivers	Recypient[a] Recipient[a]	Długość w km Length in km		Powierzchnia dorzecza w km² Drainage basin area in km²	
		ogółem total	w tym w Polsce of which in Poland	ogółem total	w tym w Polsce of which in Poland
Wisła	Morze Bałtyckie Baltic Sea	1022	1022	193960[b]	168868[b]
Odra		840	726	119074	106043
Warta	Odra	795	795	54520	54520
Bug	Narew	774	590	38712	19239[c]
Narew	Wisła	499	443	74527	53846
San	Wisła	458	457	16877	14426
Noteć	Warta	391	391	17302	17302
Wieprz	Wisła	349	349	10497	10497
Pilica	Wisła	333	333	9258	9258
Bóbr	Odra	279	276	5874	5830

Source: Mały Rocznik Statystyczny Polski 2018, Główny Urząd Statystyczny

★★ Sharing data in a structured form (e.g. a spreadsheet instead of a scanned image of a table)

An example of data shared in the structured form of a spreadsheet

	A	B	C	D	E	F
1	**The name of a river**	**Recipient**	**Total length in km**	**Length in km in Poland**	**Total basin area in km²**	**Basin area in km² in Poland**
2	Wisła	Morze Bałtyckie	1022	1022	193960	168868
3	Odra	Morze Bałtyckie	840	736	119074	106043
4	Warta	Odra	795	795	54520	54520
5	Bug	Narew	774	590	38712	19239
6	Narew	Wisła	499	443	74527	53846
7	San	Wisła	458	457	16877	14426
8	Noteć	Warta	391	391	17302	17302
9	Wieprz	Wisła	349	349	10497	10497
10	Pilica	Wisła	333	333	9258	9258
11	Bóbr	Odra	279	276	5874	5830
12						

★★★ Using open formats (e.g. CSV instead of a spreadsheet)

Example of a CSV format illustrated:

- ***in a spreadsheet table***

	A	B	C	D	E	F	G
1	The name of	Recipient	Total lengt	Length in k	Total basir	Basin area in km2 in Poland	
2	Wisła	Morze Bałtyckie	1022	1022	193960	168868	
3	Odra	Morze Bałtyckie	840	736	119074	106043	
4	Warta	Odra	795	795	54520	54520	
5	Bug	Narew	774	590	38712	19239	
6	Narew	Wisła	499	443	74527	53846	
7	San	Wisła	458	457	16877	14426	
8	Noteć	Warta	391	391	17302	17302	
9	Wieprz	Wisła	349	349	10497	10497	
10	Pilica	Wisła	333	333	9258	9258	
11	Bóbr	Odra	279	276	5874	5830	
12							

• ***as a text file*** ★★★★

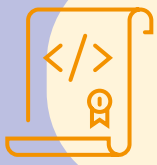
```
The name of a river; Recipient; Total length in km; Length in km in Poland; Total basin area in km2; Basin area in km2 in Poland
Wisła; Morze Bałtyckie; 1022; 1022; 193960; 168868
Odra; Morze Bałtyckie; 840; 736; 119074; 106043
Warta; Odra; 795; 795; 54520; 54520
Bug; Narew; 774; 590; 38712; 19239
Narew; Wisła; 499; 443; 74527; 53846
San; Wisła; 458; 457; 16877; 14426
Noteć; Warta; 391; 391; 17302; 17302
Wieprz; Wisła; 349; 349; 10497; 10497
Pilica; Wisła; 333; 333; 9258; 9258
Bóbr; Odra; 279; 276; 5874; 5830
```

The above examples use similar ways of recording numbers. **This is a crucial issue while preparing for data opening.**

Recommendations

Use a uniform standard while saving data. It will allow the computer to "know" which data refers to the number, which to the date, and which is the name of given categories. It will enable you to compare, analyse, as well as combine data, or even use it to create visualisations, products, and services.

Public data is not usually created by one person, and standards for data saving are rarely implemented in institutions cooperating in the preparation and sharing of data, so much work is required to "clean" data and adapt it to different needs. For this reason **we recommend using a uniformed standard from the moment of data collection**.



The Ministry of Digital Affairs has presented **detailed technical standards of data openness, which take into consideration the differences resulting from 5 levels of openness**. Before you start the data opening process, we encourage you to read the entire document.

3. Legal environment and methods of data publishing

The sharing of public data has raised doubts for many years and offices often prefer not to distribute it just in case. Below we will try to dispel these fears.

More detailed information on how to deal with, e.g. open data anonymisation can be found in two research papers prepared by the Ministry of Digital Affairs: **security** and **legal standard**.

3.1 Access to public information

Access to public information takes place according to the rules set out in the **Act of 6 September 2001 on Access to Public Information**.

According to the definition of public information issued by the Supreme Administrative Court in one of the judgments from 2002[4] and often referred to by administrative courts:

Public information is "any information generated by public authorities in a broad sense and by persons performing public functions, as well as other entities that exercise this power, or manage municipal property or State Treasury assets, within the scope of their competence. Such characteristics will also have the **information not generated** by public entities, but **referring to these entities**. Thus, public information is any information generated or referring to public authorities, as well as generated or referring to other entities performing public functions in the scope of performing tasks of public authority and management of municipal property or State Treasury property."

Before sharing public data, check if there do not occur prerequisites listed in Art. 5, Section 1 and 2 of the Act on Access to Public Information, prerequisites restricting the transfer of it to any interested person. According to this provision, sharing of public data may be denied if they constitute the entrepreneur's secret, if it

[4] Judgment of 30 October 2002. II AC 181/02 (not published).

violates the right to privacy protection or refers to other secrets provided by the law. It should be noted that according to Art. 61, Section 3 of the Constitution of the Republic of Poland, these secrets must be directly listed in other statutory provisions. Examples of such secrets may include **tax confidentiality** concerning data identifying the individual taxpayer, **statistical confidentiality** protecting the data of persons from whom the data were collected, and finally **classified information**, disclosure of which may endanger public security. In the latter case, the matter of sharing data or the inability to provide data to the applicant seems to be relatively simple. If the information is presumed to be classified, it will be provided with the appropriate security classification by an authorised person.

Also, in the case of entrepreneur's secret, pursuant to Art. 11, Section 4 **in the Act of 16 April 1993 on Combating Unfair Competition**, the entrepreneur is obliged to properly reserve certain data as an entrepreneur's secret. However, remember that **it cannot be done completely arbitrarily**.

More problems arise at the interface of the right to public information and the protection of the right to privacy. These are often individual matters which require analysis. In the famous case of disclosure of data of persons who do not perform public functions, but concluded with the office the agreements on services or delivery of goods, the **Supreme Court** clearly indicated that this form of citizen's relationship with the office determined that their right to keep personal data confidential is limited.

Remember that the protection of privacy will not be used by people whose data is provided in the context of the fact that they represent legal entities, or legal regulations directly indicate that their data is made public, e.g. **personal details of persons, who have discontinued paying local taxes**.

3.2 Re-use of public sector information

Open data can be useful for building applications, products or services. The legal framework for re-use is set out in the Act of 25 February 2016 on the Re-use of Public Sector Information. This Act implements the regulations of Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information.

Public sector information is any content or part thereof, regardless of how it is recorded, in particular in paper, electronic, audio, visual or audiovisual form, which is in the possession of obligated entities (public institutions and some companies with public property).

In practice, public sector information is based on the same subject matter as the definition of public information, the difference being that it may contain other data such as meteorological or museum materials.

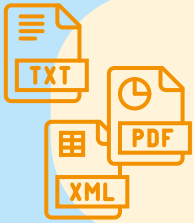
Re-use - the use of public sector information by natural persons, legal entities and organisational units without legal personality, for commercial or non-commercial purposes, which are other than the original purpose for which the information was generated.

This definition is best presented using a culinary example. If baking an apple pie is equivalent to re-using public information, e.g. a service that provides citizens with transport data, then we need products such as apples, flour and sugar, i.e. data on routes, timetables, locations of stops. We need to ensure the high quality of the ingredients and their freshness. Without meeting these conditions, no baking will succeed. These ingredients need to be combined with each other to enjoy the dessert - the transport service which operates efficiently.

Recommendations

Make sure that the data is of good quality and up-to-date. Without meeting these conditions, no service will be possible.

In the case of sharing or transferring data for re-use using ICT systems, you need to remember the obligation to use the data formats listed in the regulations of the Council of Ministers of 12 April 2012 on the National Interoperability Framework, minimum requirements for public registers, information exchange in electronic form and minimum requirements for ICT systems (Journal of Laws of 2017, item 2247). If possible, it is important to use open machine-readable formats with metadata.



Formats for data containing text, text with graphics and multimedia documents: TXT, RTF, PDF, XPS, ODT, ODS, ODP, DOC, XLS, PPT, DOCX, XLSX, PPTX, CSV. To define the information system consisting in the definition of information elements and links between them, XML is used among others.

More information on the re-use of public sector information, including allowable re-use conditions, procedures, sharing information on request, issues, calculation of fees, and also information on using potential licences, can be found in **publications of the Ministry of Digital Affairs: "Re-use of public sector information"**.

3.3 Re-use of public sector information, and the issue of personal data protection

Public sector information that is passed on to anyone interested in reusing it may contain personal information. Then re-use of such information or parts thereof will be excluded or limited. However, whenever privacy and personal data protection are at stake,

a balanced approach must be taken with regard to the equivalency of the provisions on re-use and personal data protection.

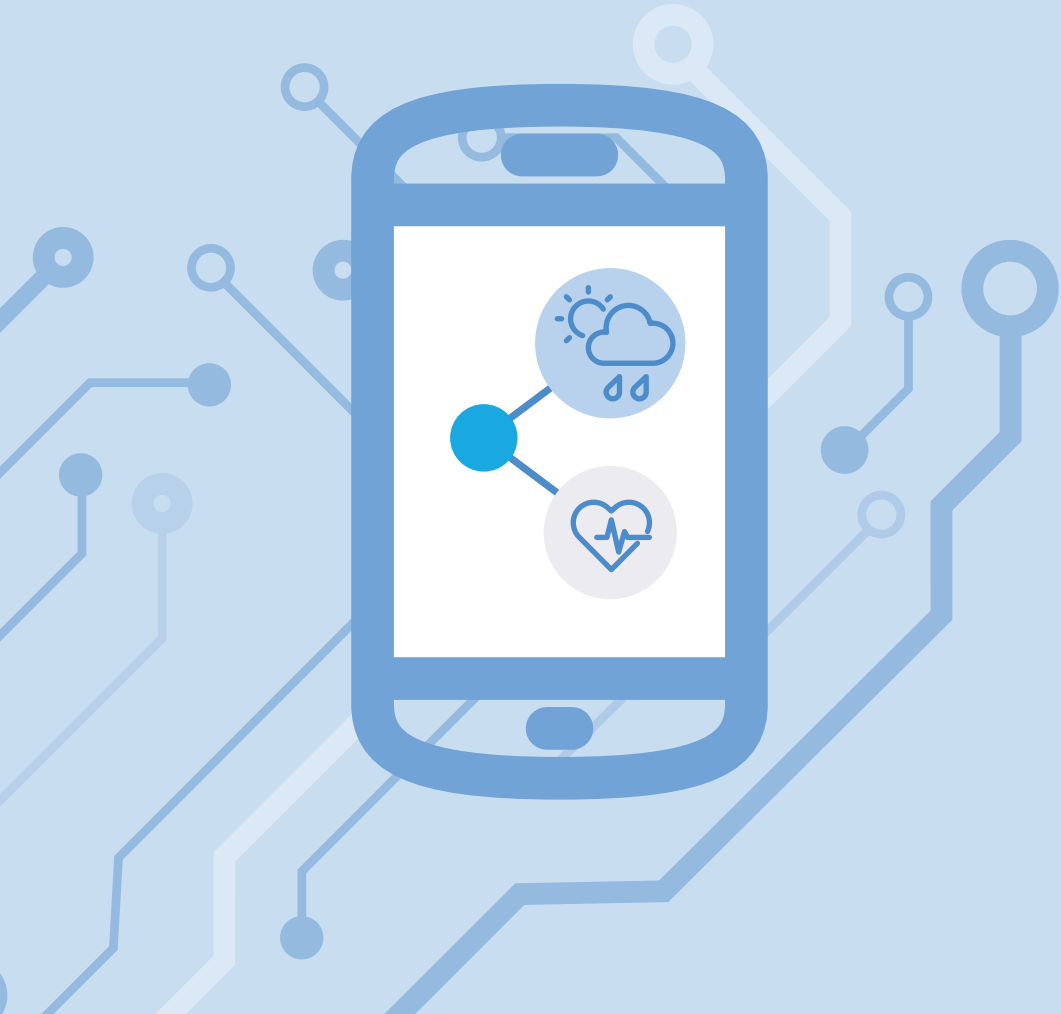
In this regard, GDPR upholds **the opinion of Working Party Art. 29 [an independent team of representatives supervising the protection of personal data in EU Member States]**, in which it was stated, among other things, that:

Further use of the already disclosed personal data for greater transparency (for example, re-publishing data by the press, or further dissemination of the originally published collection in more innovative and user-friendly ways by non-governmental organisations) may also be desirable. Whether this re-publication and re-use is possible will also depend on the outcome of the balancing test, which should take into account, among other things, the nature of the information and the effect of re-publication and re-use for individuals.

Therefore, GDPR does not, in principle, limit the re-use of public sector information in the context of personal data that does not only concern the private sphere. However, every effort should be made to assess whether it is acceptable to change the purpose for which the data was collected in any given case.

If in doubt, as in the case of sharing data on the basis of the provisions on access to public information, it is best to consult with the person responsible for the personal data protection policy at the office.

4. How to share public data?



4.1 Open data portals

Recommendations

Publish data online. This will guarantee that the potential of open data can be used.

Regarding the mode of data sharing on request, we refer to chapter 3.2 of the text on the re-use of public sector information. It should be mentioned in the proposal provided for in Art. 21 Section 2 of the Act on the re-use of public sector information, i.e. the implementation of the proposal, by allowing for a period of no longer than 12 months to re-use public sector information collected and stored in the ICT system of the obliged entity in a permanent, direct way and in a real-time manner.

Some of the municipal offices, such as **Gdynia Municipal Office** or **Warsaw Municipal Office**, create their own open data portals containing data generated by the municipal office and its units or relating to the tasks they perform. Open data portals make it easier - especially at the urban level - to find specific information because, by definition, they constitute a repository of the majority of datasets that can be reusable in a given city hall or its organisational units.

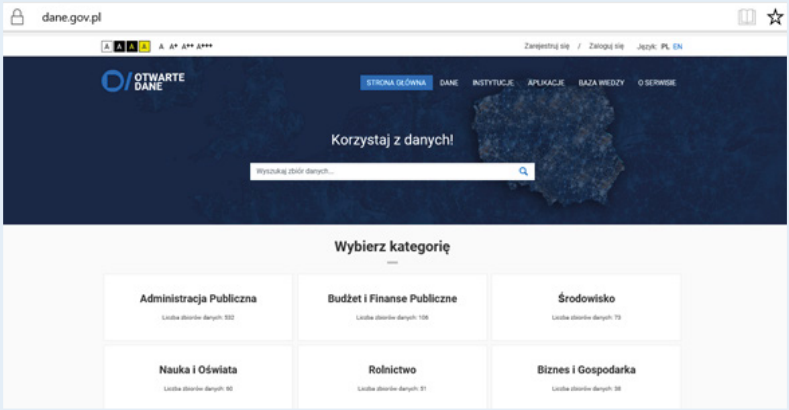
The creation and maintenance of such a portal is, however, associated with the significant cost of maintaining your own team who are developing the portal, or the risk of becoming dependent on a commercial service provider. Modernisation and complementation are a challenge when it is not managed by an interdepartmental team in the office which deals with open data. Without a coherent vision and constant data supply, the portal loses its attractiveness, and as a result, the motivation of office workers to prepare data and publish them decreases.

An example of an impressive portal is the **portal containing urban data in London**. It uses uncomplicated tools (CKAN and WordPress), but it is the effect of the work of a dozen people whose responsibilities are also to encourage the use of the portal or to cooperate with the organisations involved in the use of public data.

Publishing data in Public Information Bulletins is not a solution either. They are not usually adapted to function as data repositories, especially if we want to **share data via API**.

Recommendations

Before you start creating your own data repositories, consider the possibility of publishing data in the Central Repository for Public Information (CRIP), or on the portal: dane.gov.pl



Some institutions operating under central administration are already obliged to transfer certain groups of data to the CRIP. The Act on Access to Public Information, which regulates the means of conducting CRIP, also provides the possibility of publishing other resources, for example, those transferred by local government units, which already happens.

A very important element which accompanies the sharing of data on open data portals, including CRIP, is a reliable description of the information that is published, i.e. **metadata**.

Metadata is the information about data - structured information used to describe information resources or information objects, providing detailed data on asset attributes or information objects to facilitate their finding, identification and management.

Regulation of the Council of Ministers of 12 March 2014 in the case of Central Repository for Public Information determines what information must be included in the description of the sets shared on this portal. This minimum set should also be used on any open data portal.

1. **Metadata which describe the information resource entered by the provider includes:**
   a) title - the name of the information resource that allows its identification;
   a) keywords - a set of words or phrases that briefly describe the information resource;
   a) description - a summary or short description of the content of the information resource;
   a) update frequency - information on the update frequency of the information resource or metadata.
2. The provider also introduces the URL address of the website or its subpages, within which these information resources were shared.

Whenever data is published on open data portals, **we also recommend adding a short description explaining the methods of data collection,** i.e. what methodology was adopted and whether the original data (indicator cards) were collected, and if so, from what source.

4.2 Data visualisation

Presenting open data on charts or in the form of attractive visualisations goes beyond the scope of data opening - that belongs to the stage of their use.

However, data visualisation can significantly facilitate data analysis. Sometimes it is enough to present the data on the map, thanks to which we will easily notice places with the highest intensity of a given phenomenon. This is an example of how the above-mentioned **national map of threats**, **map of listed buildings** or **map of road hazards work**. Other data, especially numerical data, looks good on charts. Thanks to this, it is easier to imagine the scale of a phenomenon or follow changes in time. This can be done in the way the Ministry of Finance presents **data on the implementation of the State budget**, or Gdańsk presenting the **dynamics of various phenomena over several years**.

However, remember that the type of visualisation must match the nature of the data presented. If we present data on the expenditure of a city in a series of years, we will use a line chart. However, this will not work if we want to combine the most important expenditure categories in a given year on a single chart. You can read more about the principles and tools for data visualisation in the publication **"Mów do rzeczy"** from Foundation "Shipyard" - Centre for Social Innovation and Research, and about data presentation in the **Central Statistical Office publication, which discusses cartographic standards of statistical data presentation**.

Presenting data in graphic form can also help to reach a wider range of receivers, especially those not prepared to analyse data in spreadsheets or CSV files. As an example, the visualisation of

the **budget of Gdańsk** can be used so that every taxpayer can better understand what the city spends the most on and how the taxpayer "contributes" to the city's finances.

There are many products on the market that allow us to visualise large data resources in a simple way, starting with the most popular spreadsheets, through dedicated commercial and **non-commercial** solutions that work well in both the internal environment and in online publications.

At the stage of creating the visualisation we can frequently detect errors that arise during the collection or preparation of data. That is why it is worth visualising data even for internal purposes, to make sure that the data provided is correct.

5. How to organise work in an office

Success in opening public data is possible only thanks to a comprehensive approach. That is why it is worth creating strategies, data opening policies at the level of each office, drawing from proven patterns in other institutions

Recommendations

Design the proces of opening data in your own government office in such a way as to fully use the potential of the given unit.

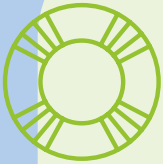
The "silo" form of Polish administration also affects the area of public data. Lack of cooperation between departments, imposing additional duties, insufficient knowledge or difficult communication within the office mean that in place of a consistent policy of opening public data, individual and incidental solutions appear which do not fulfil the objectives of opening public data.

Main problems in the proces of opening public data:

- lack of a coherent policy of data opening at the entire office level;
- rigid, hierarchical structure of the office;
- lack of knowledge about legal regulations, both regarding legislative area and contractual provisions;
- lack of knowledge and understanding among officers about the importance of the role of data opening;
- lack of a person coordinating and inspiring the process of data opening;
- lack of an interdisciplinary team in charge of the process of data opening.

5.1 Development and implementation of the data opening process. The role of the person managing the office and the team leader.

The answer to all above-mentioned problems is the implementation of internal openness policies which, at the same time, appoint horizontal management structures for the process of opening public data. Examples of such solutions can be found in some Polish cities. Openness policies were introduced, among others, in **Gdańsk**, **Słupsk**, **Łódź** and **Starachowice**.



5 steps for effective implementation of the openness policy

1. Making decisions on a political level to inspire the process of data opening.
2. Choosing a leader and defining his/her tasks.
3. Establishing a team, and the commitment of other employees to cooperate.
4. Developing the openness policy (defining data catalogues, scheduling of sharing, defining standards, and people responsible for implementation).
5. Ensuring implementation and evaluation of the openness policy.

An indispensable element of openness policy introduction in an office is to gain support among a close group of decision-makers **with the person managing the given unit at the head**. This person has to take responsibility for presenting solutions that comprehensively regulate the area of data opening and, at the same time, affect the change in the employee management system.

Every team needs a leader. The leader coordinating the work of the data opening team should be a person working in the unit coordinating the work of the entire institution, interested in the subject of data opening, and who understands the needs of residents, non-governmental organisations and business.

The person, on whom the success of the data opening process depends in the office, should implement the policy in the office and coordinate the work of the team. His/her involvement and ability to mobilise the team determines the success of the process.

Below you will find **a few tips** to help you[5].

1. **Show your vision with clear examples of the benefits that open data will bring**. Present your problem and how open data can affect its solution - not only from one dataset, but as a result of combining data in different departments or institutions. It will help other people working in the office to identify with the data opening process.

2. **Build understanding for data opening among people managing departments**. They will not be directly responsible for the process, but their acceptance is crucial for the motivation of employees. Encourage them to take part in a short training course to show the potential benefits for their unit and the entire office.

3. **Identify people who share the vision of public data opening**. The management's support is important for ensuring process stability, but drawing in "reformers" will allow innovation and continuous development.

4. **Establish communication channels dedicated to gathering comments and opinions from people** who will not be involved

[5] Based on The Open Data Institute. Open Data In Government: How To Bring About Change. In the context of management of the data opening process in government offices we also recommend The Open Data Institute. Lessons On Supporting Public Sector Open Data Leadership.

on a daily basis in the data opening process, but who take part in creating, developing and using data within the scope of their duties. Thanks to this, you will build support for the openness policy among the larger group.

5. **Show that small successes are also possible**. Determine to start with opening small datasets or carry out a pilot showing how they can be used to the benefit of the office. Remember, however, that they must be part of a comprehensive and long-term openness policy. Combining quick victories with long-term goals can help maintain the dynamics of change.

6. **Be flexible. React to ideas and needs of different departments and other teams**. Express your long-term vision of open data, but be careful from the beginning to avoid being blocked by the fixed plan to implement an open data policy. Being flexible helps you to adjust to changing technological tools, society needs, and changes in the functioning of the administration.

7. **Make sure that within the Open Data Team there are people responsible for change management** - not only at the level of specific activities, such as building a portal, publishing datasets and animating them again to be used. People in the team should look for synergy and, in the process of managing change, they should also take into account other factors, seemingly unrelated to the opening of data.

8. **Get external support for your actions in a business, non-governmental or academic environment** to drive a steady demand for open data. External support can help maintain political will for the process of data opening and be a source of continuous learning and dialogue.

9. **Make sure that officers have the possibility to learn**. They can participate in training, conferences or even temporary work with a data analyst or a person dealing with technologies. This will help in further improving the open data policy.

5.2 Appointment of the team

Implemented openness policies should introduce a **horizontal management structure** of the opening data process. It is necessary to appoint a team composed of people whose skills and knowledge will ensure the full implementation of data opening potential.

While appointing the team, you should be sure to invite **people responsible for**:

- **development of strategies based on the collection and use of data** (e.g. strategies for solving social issues);
- **implementation of technological tools**;
- **protection of personal data and legal issues**;
- **processing large amounts of data** (e.g. accounting data).

In Gdańsk, the team includes, among others, the Information Security Administrator, an employee of the Department of Promotion and Social Communication, the Director of the Informational Technology Department and the coordinator on behalf of the mayor's office. In Starachowice, the team also includes a legal consultant and an inspector from the Financial Department.

The team should be responsible for at least the following tasks:

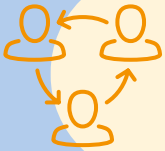
1. Preparation, coordination, monitoring, and evaluation of data opening strategies.
2. Developing standards of data openness (or choosing the standards prepared by another institution, e.g. the Ministry of Digital Affairs) and their implementation in the office.
3. Inventory and analysis of available public data.
4. Developing and testing the IT environment which will be used for processing the data (e.g. API).
5. Drawing up schedules for sharing public data.

6. Cooperation with non-governmental organisations, business sector representatives, and other institutions that have an interest in opening and using data.

An important part of the team's work should be regular (at least once a year) **drawing up of the public data availability report for the previous period**. The report should also include an action plan for the next period designed to improve access to online public data.

The key to an effective data opening process is introducing planned changes to the employees, keeping in mind their experience in everyday work with data, as well as the fact that they will most likely have more responsibilities related to the implementation of the policy objectives.

5.3 Development of openness policy



Similar to the case of other documents prepared by public institutions, we recommend the openness policy should be adopted only after an extensive public consultation with both the office employees and the residents.

The open data policy is not aimed at the people directly responsible for its implementation. Due to its comprehensive character, the policy is also aimed at other office employees as well as individuals and institutions who may potentially use public data. The openness policy is also of educational nature. Therefore, it should include the following items[6]:

1. **The definition of open data** and the objectives of data opening, as well as the explanation of their importance to the office and the reasons for the policy implementation.
2. **General statement of principles** which should guide the publication and re-use of open data (including drawing up the rules for publishing data online).

[6] Based on: https://theodi.org/article/how-to-write-a-good-open- data-policy/.

3. **The scope of data categories** collected by the office and whether the data can be made available for free re-use.
4. **Reference to all relevant legal acts, policies, or other guidelines**, which also apply in relations with citizens and other entities interested in the use of data, including the acts or policies regarding privacy issues.

A key element which will guarantee the harmonisation of the process of data opening is, alongside the appointment of a team, **obliging employees in charge of all departments of a given institution to cooperate with the team and to actively participate in this process**. Furthermore, it is crucial to gather knowledge about the data held by the office.

Extract of Regulation No. 306/14 of the Mayor of the City of Gdańsk, dated March 12, 2014.

1. It obliges directors of organisational units to create an Open Data catalogue.
2. The catalogue should include:
   a) a list of public datasets created by the organisational unit;
   b) data within the set should be grouped into data that is already available and data that is planned to be shared.
3. Lists of sets and data contained in them should be created by organisational units under the supervision of, and in cooperation with, the Team.
4. The directors of organisational units are obliged to cooperate permanently with the Team in the implementation and development of the openness policy strategy through:
   a) posting of employees to work with the Team;
   b) undertaking an analysis of available public data;
   c) preparing appropriate public data to be published;
   d) gradual publishing of new datasets in accordance with a schedule developed along with the Team.

Keep in mind that **the way information concerning data catalogues is presented is important and should be readable**. It is worth dividing them into categories (e.g. health, finances, demography), marking them with keywords, as well as indicating which ones can be downloaded by the API.

A good openness policy should contain its **implementation schedule** and it should indicate clearly r**eporting obligations** resulting from it.

When creating a schedule for opening data, consider the following rules[7]:

1. **Define priorities**. Begin by opening data that is of far-reaching importance to citizens and institutions: data needed to develop a legal act, strategy, policy or to facilitate the implementation of new regulations in the office. Then, open data which is very popular among citizens, organisations and businesses. Do not limit yourself to things that people interacting with the office want. In cooperation with other employees of the office, decide which data will help improve the functioning of the institution, or improve the quality of life of residents.
2. **Pay attention to the costs**. Sometimes the cost of data opening is high due to their complex structure. Keep that in mind while preparing the schedule. Perhaps it will be necessary to gain additional funding, which means it may be worth postponing data opening to achieve a better effect later on. That is why it is worth starting from data that is already published and of which the adaptation to standards of openness is not complicated.

The policy has to contain a precise indication of who will review the shared data and how. Such an obligation should belong to a team that cooperates with the managers of departments and other units, and who will set the specific dates for checking correctness (currentness, and compliance with standards) of data, also taking into account the frequency of their download and ways of being

[7] Based on https://sunlightfoundation.com/opendataguidelines/.

used, the changing technological environment, and feedback from users. The role of a leader is to monitor the implementation of the openness policy and to respect the implementation schedule on a daily basis. It is a good practice to define data standards both relating to the stage of their creation, further development, and publishing. It is worth striving to ensure that standards are also consistent across many offices if they apply to the same data. The recommendations contained in the **technical standard developed by the Ministry of Digital Affairs** will help in ensuring data consistency.

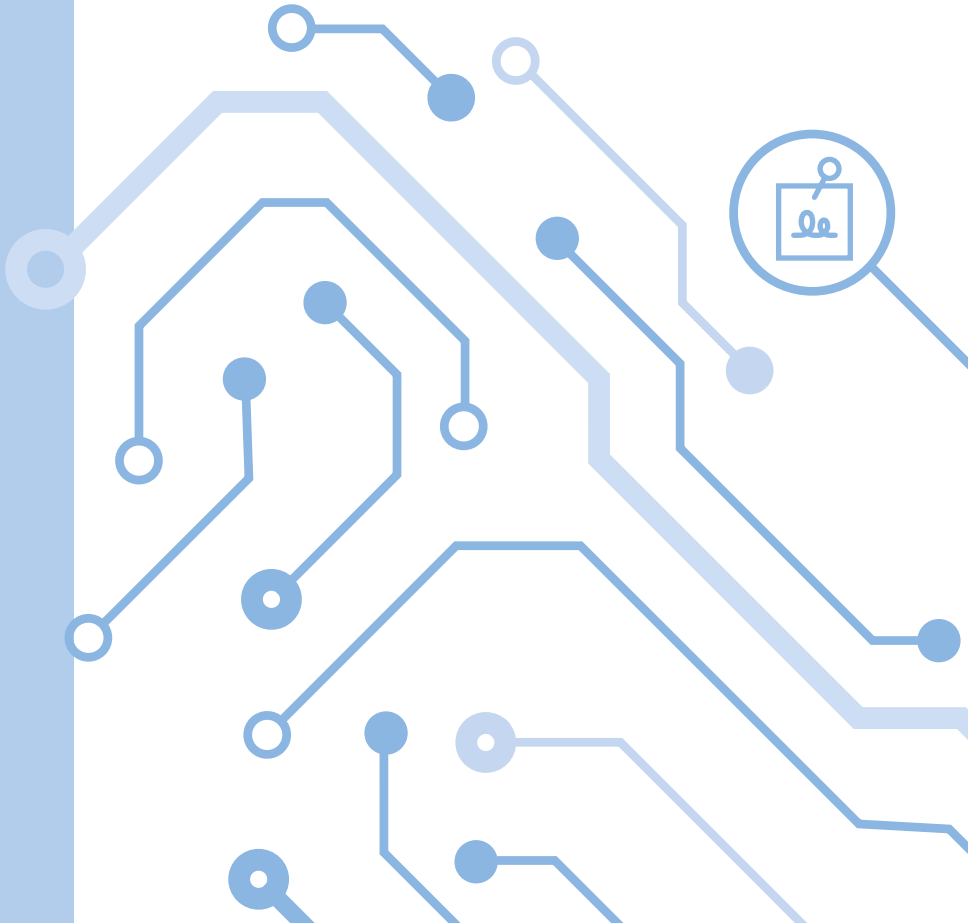
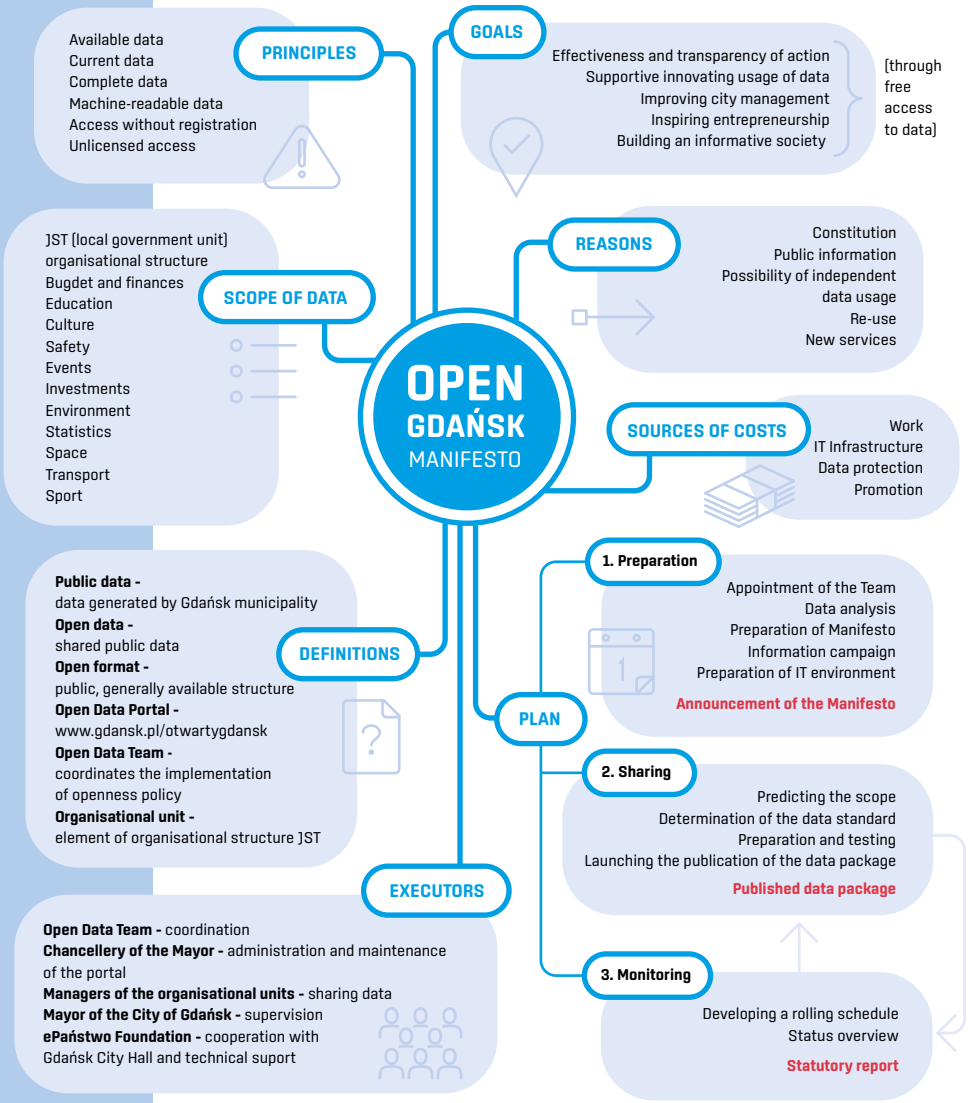


Diagram of the Policy of Openness based on the example of the City of Gdansk[8]



[8] Attachment No. 2 to the Regulation No. 306/14 of the Mayor of the City of Gdańsk, dated March 12, 2014. Available at: http://otwartygdansk.pl/wp-content/uploads/2016/07/ PMG_2014_6_306_zal013.pdf.

6. Openness policy implementation schedule. Dataset opening

In the earlier parts of this manual we presented information concerning the creation of an institutional and strategic environment for the data opening process. Below you will find the schedule for opening specific datasets.

1. Identify your datasets

Honestly, there is no need to share all public data at once. Start with data that is the easiest to share, which means data meeting certain standards or those which can be easily adjusted (see chapter 2. What does open data mean). At this moment, it may be worth organising training for people working in the office to unify concepts, diagnose internal needs of officers and make sure that data can be shared. Use the knowledge base available at the portal **dane.gov.pl**.

We offer tactics for action that can be treated as consecutive elements of the process or as separate activities:

1.1 Inventory databases in all units

There are different styles of running databases related to the daily work of the office. You need to carry out an inquiry among them to find out what information is collected. Very often the same units or departments do not know what kind of databases other units have. A good example is **an inquiry conducted during the implementation of the Program**. Thanks to the inquiry you will find out what kind of data are open.

1.2 Analyse motions and queries concerning access to public information and the re-use of public sector

Citizens surely submit motions for data sharing in your office. The mode of access to public information in particular has become quite popular and allows citizens to receive public information that is not available in friendly formats. Repeated queries from citizens

signal that it is worth sharing information in a friendly way. Other channels through which citizens submit requests concerning data, information, files or services can also be analysed. On the basis of the above analysis, choose a few of the most important databases to be opened in the first place.

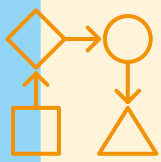
1.3 Check legal conditions

Check provisions of law regulating the data registers which you have identified. If you purchase or share a song, please remember appropriate copyright protection. It is good practice to make it available under a free licence. More information about legal issues can be found in a legal standard.

2. Ask within your environment about needs

2.1 Identify the most important stakeholders

Considerable financial and time expenditure can be devoted to providing access to certain databases for which there is no social demand. It may cause resistance and doubts within the office concerning the sharing of other data. Therefore, it is worth creating the opportunity to get to know the needs of the office environment and look for allies of the project. Who usually asks for data or information? Start with the office partners who usually request information or with whom your office already works, for example, in the form of assigning tasks. Also groups such as businesses, social organisations or scientists use data in their work, therefore they know very well what data they need and what is lacking. It is worth being present at industry meetings to observe in what cases there is currently a need for information, analysis or datasets.



What can interdisciplinary cooperation and data exchange between public administration units bring?

Inspired by Prof Jonathan Shepherd, who noted that ¾ of the victims of violence administered in hospitals do not report the incident to the police, the city of Cardiff in Wales created a system for exchanging anonymised information between hospital admission rooms and the local authority and the police. In hospitals, victims were asked four short questions, and the responses after anonymisation were shared with the police. Police, along with representatives of the authorities, analysed cases collectively and introduced changes in places where crimes most frequently occurred. This led to savings of $6.6 million and reduced the level of violent crime by 42%. The model operates in various forms throughout the United Kingdom; it has also transferred to the United States, Australia, the Netherlands (all hospitals use this model) and South Africa.

Source: Wales cuts reported violence by 42% in its capital with data sharing model, apolitical.co; Cutting crime: surgeon's research that helps stop street violence, theguardian.com.

2.2 Conduct individual interviews with the most important stakeholders

Conducting interviews with the most important subjects, who are interested in the data, will allow you to obtain detailed information concerning their needs in terms of opening the data as well as potential consequences and profits connected with satisfying their needs. If you conduct interviews only with a specific group, you will not get the full picture. It is worth diversifying the test group in order to recognise different needs within one area. Use your knowledge to generate one or more case study of the data opening

How was it done in New York?

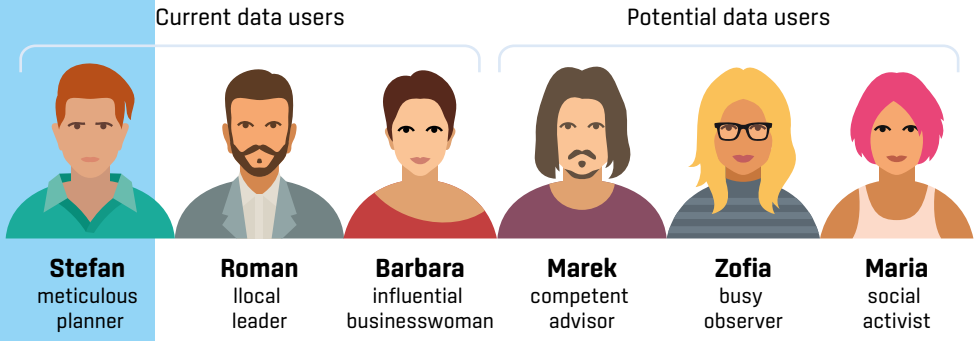
It was determined that during the interviews the same number of respondents would be tested in terms of using public data with the following frequencies:

- **low:** never used public data and never heard about it;
- **medium:** browsed the data and/or asked about it, downloaded from 1 to 5 sets;
- **high:** have a broad knowledge in terms of datasets and how often they use them.

Source: „Understanding the Users of Open Data: Research Findings".

For example, in New York, as a result of interviews, the following types of users of Open Data were created.

Meet data users from New York

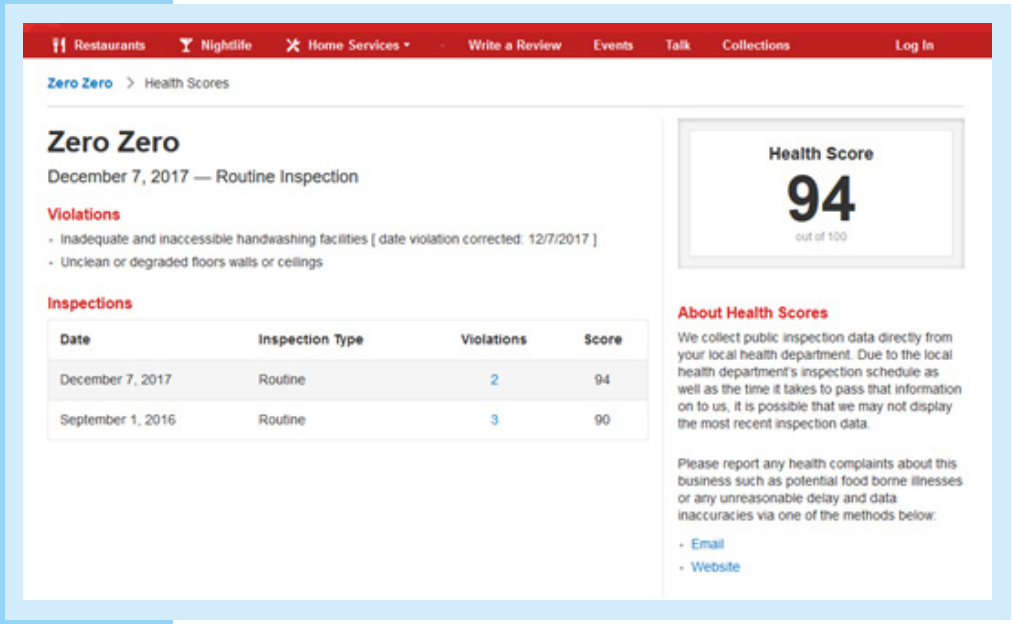


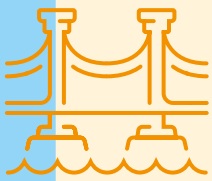
Source: „Understanding the Users of Open Data: Research Findings".

2.3 Choose a detailed topic

On the basis of previously taken steps, choose one of the cases of data usage from a specific sphere, for instance on the basis of an identified need from one of the office's partners. Consider your own resources and capabilities as well as the needs of the environment. **At the beginning, it's worth choosing something simpler to offset the risks, and practise the opening data cycle itself.** The key is to focus and limit yourself to one case, one need at a time. It is worth maintaining close contact with a given organisation or entity whose needs have been chosen for implementation. Their comments and feedback are necessary to open a dataset that responds to the needs.

Cooperation with Business: Sanitary Ranking in San Francisco





San Francisco Department of Public Health shared data from sanitary inspections in restaurants. The yelp.com portal used these data to create an additional evaluation criterion for restaurants so that residents could learn about the level of hygiene in a given restaurant. In addition, inspectors used a new way to choose places for sanitary inspections. They receive special reports from the portal that draw attention to restaurants in which clients complain about food poisoning and low levels of hygiene. The Harvard Business Review study showed that the publication of sanitary control assessments leads to the situation in which low-rated restaurants tend to increase their level of hygiene and get better results in subsequent inspections.

More: **Cities take to Yelp to tell their residents where not to eat**, apolitical.co.

3. Check and standardize your data

See what kind of data you have in your database that you want to open. Check that there are no errors in it, that all fields have the correct values, whether there are any "holes" in statements, typos and other errors that reduce the usefulness of the entire database. Set one standard for your data (one way to save the date, address, location, value). To ensure consistency of standards, use the **Ministry of Digital Affairs standards**, also in the scope of **API**. It is also worth using feedback from potential users of the datasets, which opens or consults the usefulness of the database with experts.

Make sure that you provide data in a format suitable for partners who have presented their needs.

4. Enter context - metadata

Do not publish the database itself. Write down what's in it, add understandable keywords (rather: "car" than "vehicle"), add conditions for the use of this data, the scope of data, and their timeliness (date of creation, date of last modification, frequency of updates). Such information will help users to get an idea of exactly what the set represents. Use the **guidelines of the Ministry of Digital Affairs**, also in the scope of **API**.

It is also worth showing the results of the diagnosis of needs for a given database, collected from the office environment. It will allow the context of data usage to be shown and may increase the use of the database by less advanced users. Invite data users to share real data usage histories for a certain purpose.

5. Choose the open publication format

Use, for example, CSV or XML - remember that the data must be machine-readable, so scans or "manually" created tables in a text document are not the best solution.Use the **guidelines of the Ministry of Digital Affairs**, also in the scope of **API**.

6. Publish the data

On the wave of enthusiasm, some municipal offices in Poland created their own dedicated open data portals. However, the experience of some local governments shows that maintaining your own services is time-consuming and expensive. The solution to this situation for all public entities may be to place data on dane.gov.pl.

You can also place datasets on your BIP, provided it has the appropriate technical parameters and allows you to easily search for data. It is important not to limit the access to the page and to make it readable for the users.

7. Make sure that the data is available

To ensure that data is up-to-date for users using machine-based data access, it's best to attach to the shared data **API (Application Programming Interfaces)**, which will mean that as soon as you make a change, you will immediately update your data and those of people using them. This is very important for data that changes frequently over time or is simply made available in real time (current air pollution, water status information, traffic obstruction messages, etc.). Take advantage of the standards developed by the Ministry of Digitisation in the field of **API**.

8. Inform the world

Data has meaning if it is used. Your environment, both close and further away, should learn about new datasets.

If you have established relationships with the most interested parties based on previous steps, they are the "tube" that will tell the world that we have new, interesting datasets.

It is worth contacting technology bloggers, universities, scientific circles and non-governmental organisations. Inform them that you have data and they can use it. Such a broad approach can provide additional feedback. Listening to these voices may result in proposals for interesting solutions that will increase the usability of the data.

If you want to promote open data or identify interesting needs of the office environment, which can only be satisfied by opening datasets, you can organise a hackathon/datathon event.

This is a type of programming or analytical marathon that takes place most often over the weekend. During such an event, enthusiasts of various specialties try to develop innovative ways to solve a problem. Thanks to that they can provide many interesting solutions based on public data from various sources and on **open source software**. Such software makes it possible to customise a technological tool that has already been created by someone else (for instance **https://www.opentreemap.org/**).

Such events are worth organising in industry partnerships with other public institutions and businesses or non-governmental organisations. If the reward is to implement the tool in your office, or support in independent implementation, the sense of influence on the reality of the winning team may be sufficient motivation.

An important element of the hackathon/datathon is to define tasks/challenges that cannot be too general. At this point, the specific needs that have been collected from the environment will be very useful at the initial stage of the process of opening your bases.

National Museum in Warsaw organised in May 2018: HackArt / Hackathon of the National Museum in Warsaw to find new ways to use their sets (treated as collections in the public domain). The main prize was, among others, the implementation of the best

projects with the help of partner companies. In order to avoid the participants creating solutions to non-existent problems, the organisers presented five challenges that are related to the everyday work of the museum, e.g. how to help those interested in exploring collections, both virtual and located in the museum building?

In June 2018, two teams joined the community **Code for Poland** and they have been developing the project with the support of a museum employee. Together, they aim to implement, for example, a chatbot in the style of Stańczyk who answers simple questions about the work of the museum and relieves employees.

More on: **https://hackathon.mnw.art.pl/**

9. Update the data

Open data needs updating so that it can be re-used. Set the frequency of the update, let us know about it, and stick to the plan.

10. Get back to the first point

Individual steps are worth going through for each dataset being opened. This protects against the opening of low-use datasets, both when it comes to technical aspects, and in relation to the needs of the social environment of your office.

This manual was created in cooperation with employees of the Department of Open Data and Development of the Competences of the Ministry of Digital Affairs including the following: Agata Miazga, Aneta Cesarek, Anna Gos, Anna Bramska, Magdalena Dąbrowska, Karolina Kościerzyńska, Sylwia Pichlak- Pawlak, Dominik Sybilski, Iwona Szelenbaum, Jacek Żuchowicz, with the ePaństwo Foundation. The workshop, which took place on June 11, 2018 at the Gdańsk IT Centre, organised by the Ministry of Digitalisation, the City of Gdańsk and the ePaństwo Foundation, with the participation of the representatives of the City of Gdynia, served as an exchange of experience between local government and government administration in the field of opening public data. It was preceded by individual interviews with practitioners of data opening in Gdańsk. Centrum Cyfrowe was also consulted regarding the content of the manual.

The manual was developed as part of the project: **"Open data - access, standard, education"**.

If you have used the advice contained in the manual, if you need additional information or if you want to share your opinion about it, write to us at: **otwartedane@mc.gov.pl**.

Fundusze Europejskie Polska Cyfrowa
Ministerstwo Cyfryzacji
OTWARTE DANE
Unia Europejska Europejski Fundusz Rozwoju Regionalnego