Analytical Report n5

Analytical Report 5: Barriers in working with Open Data

This study has been prepared by Capgemini Invent as part of the European Data Portal. The European Data Portal is an initiative of the European Commission, implemented with the support of a consortium[i] led by Capgemini Invent, including Intrasoft International, Fraunhofer Fokus, con.terra, Sogeti, 52North, Time.Lex, the Lisbon Council, and the University of Southampton. The Publications Office of the European Union is responsible for contract management of the European Data Portal.

For more information about this paper, please contact:

European Commission
Directorate General for Communications Networks, Content and Technology
Unit G.1 Data Policy and Innovation
Daniele Rizzi – Policy Officer
Email: daniele.rizzi@ec.europa.eu

European Data Portal
Gianfranco Cecconi, European Data Portal Lead
Email: gianfranco.cecconi@capgemini.com

Written by:	**Reviewed and with contributions by:**
Jorn Berends	Thore Fechner
Wendy Carrara	Marc Kleemann
Heleen Vollers	

DISCLAIMER

---

[i] At the time this report was first issued the consortium consisted of: Capgemini Invent, Intrasoft International, Fraunhofer Fokus, con.terra, Sogeti, the Open Data Institute, Time.Lex, and the University of Southampton.

Executive Summary

Open Data can be the new raw material. Opening up Public Sector Information has the potential to offer fuel for companies to create value by making new apps and services, for public administrations to become more efficient and for the society at large to become safer and more sustainable. Owners of Open Data portals are developing and improving their portals, and these portals are increasingly backed by solid policies on the countries' digital agenda, or more specifically, Open Data. The majority of the EU28+ countries have successfully developed a basic approach to address Open Data, but are at the same time struggling with a number of barriers, hindering them to move forward with Open Data. At the same time, recent research suggests that such barriers also exist at companies that wish to make use of Open Data, restricting them to fully benefit from the potential of Open Data.

Building on a three-pillar approach, this report brings together barriers encountered by both data publishers and re-users of Open Data.

- An analysis of recent reports on the topic of the publication of Open Data, such as the 2015 and 2016 Open Data Maturity in Europe reports and the European Data Portal (EDP) Sustainability report. These findings were complemented by barriers mentioned when the EDP interacts bilaterally with Member States.

- To delve into the specific barriers faced by re-users, the results of the EDP Re-using Open Data report were analysed.

- A broader scoped desk research was conducted. Research also focused on sentiment analysis and interviews conducted with the Open Data community during conference and workshops.

The barriers discovered during the research appear at both the side of data publishers and re-users of Open Data, and are often related and dependent on each other. Portal owners suffering from a lack of technical knowledge on metadata quality on the one hand requires portal owners to invest more time and resources in the publication process, but it can also lead to inappropriate metadata being published, leading to difficulties among re-users to find the data they are looking for. Barriers are found within the political, organisational, legal, technical and financial domain, and awareness on both the availability of Open Data and the specific needs of users of Open Data leaves room for improvement. Specific barriers are found in the geospatial data domain, which uses different standards as compared to other types of Open Data. Working with geospatial data requires thorough and specific technical knowledge, different from the technical knowledge needed to work with 'usual' data.

Barriers are related and dependent on each other

Across Europe, both data publishers and re-users of Open Data have found ways to cope with these barriers. For instance, more and more portal owners are using the DCAT-AP profile, resulting in more standardisation and a better accessibility of data. Portal owners are organising events to drive user communities, national guidelines for the publication of data are put in place and users have developed work-arounds to use particular datasets. Unfortunately, solutions are not always shared across Europe, and solutions are sometimes ad-hoc, stand-alone, or only temporary. An important means to get insights in each other's needs and to find sustainable solutions to address barriers is to open up

the dialogue between data publishers and re-users. Still, it is important to realise that there is no single solution that will solve all the challenges. Instead, policies need to address a combination of interwoven solutions to the different types of interdependent challenges.

Table of Contents

1   Introduction

"Data should be able to flow freely between locations, across borders and within a single data space. In Europe, data flow and data access are often held up by localisation rules or other technical and legal barriers. If we want our data economy to produce growth and jobs, data needs to be used. But to be used, it also needs to be available and analysed." – Andrus Ansip, Vice-President for the Digital Single Market [1].

The impact of Open Data is undeniable. The digital economy revolves around Data, with (Open) Data being the new raw material. In this context, Open Data – as defined by the Open Definition - refers to the information collected, produced or paid for by public bodies and can be freely used, modified, and shared by anyone for any purpose. Furthermore, data must be available under an open licence and preferably in machine readable format. This definition can be narrowed down to Public Sector Information (PSI), or Open (Government) Data, that is to say data collected and published by the public sector.

Studies measuring the impact of Open Data initiatives underline the importance of Open Data for economic growth, to drive monetary benefits and foster transparency. Open Data not only drives economic growth, but also generates benefits for the society at large, such as reduced traffic congestion and more efficient energy use. It may even help to save lives. In emergency situations where every minute counts, Open Data can improve an effective response by analysing where to place equipment and station personnel. It was estimated that Open Data has the potential of saving 1,425 lives a year across Europe, equalling 5.5% of the European road fatalities. In terms of macroeconomic benefits in the European Union Member States plus Iceland, Liechtenstein, Norway and Switzerland (commonly referred to as the EU28+), the market size of Open Data is expected to increase by 26,8% from 2017 to 2020 to a value of 75.7 bn EUR in 2020[2]. The forecasted number of direct Open Data jobs is expected to rise from 80,000 in 2017 to nearly 100,000 jobs by 2020. Thanks to the positive economic effect on innovation and the development of numerous tools to increase efficiency, not only the private sector, but also the public sector is expected to experience an increased level of cost savings through the re-use of Open Data to a total of 1.7 bn EUR in the EU28+ by 2020.
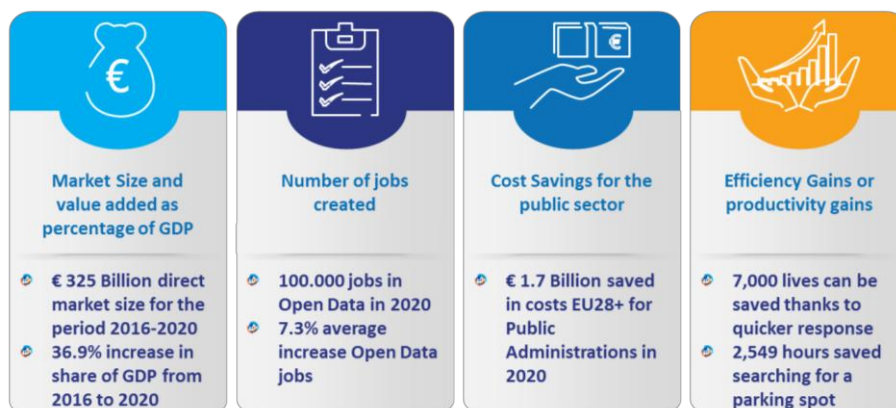

Figure 1: The economic value of Open Data

---

[1] European Commission, 2017, Commission outlines next steps towards a European data economy
[2] European Data portal, 2015, Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources

More people, organisations and countries are aware of Open Data and the promises it holds. This results in citizens and businesses using and requesting more data, public administrations opening up more data and more national governments developing a policy framework underpinning the release of Open Data[3]. The Open Data Barometer report indicates that 55% of the countries had an Open Data initiative in place in 2015, while the year before 50% of the countries taking part in the survey had an Open Data initiative[4].

These findings are also confirmed in the latest report on Open Data Maturity in Europe 2016[5]. This report states that in 2016, the EU28+ countries completed over 55% of their Open Data journey showing that, by 2016, a majority of the EU28+ countries have successfully developed a basic approach to address Open Data. This same report underlines the fact that there are barriers which impede Open Data to move forward. A report on the re-use of Open Data[6] published in January 2017 further stressed that on the user-side, a number of barriers prevent companies from efficiently re-using Open Data and leveraging the potential it holds. As illustrated by the quote of Vice-President Ansip in the beginning of this chapter, these barriers thereby prevent society to reap the full benefits of Open Data, and have a negative impact on both the publication and re-use of Open.

To identify and delve into the particular barriers faced when working with Open Data, the approach of this study was built upon three pillars.
- First, in order to better understand the barriers faced by data publishers, the results of both the 2015 and 2016 Open Data Maturity in Europe reports were analysed as well as the European Data Portal (EDP) Sustainability report[7] published early 2017. These findings were complemented by barriers mentioned when the EDP interacts bilaterally with Member States. Since the EDP harvests the metadata of Public Sector Information made available by European countries, it is in direct contact with data providers and publishers in more than 34 countries. In addition, the EDP team also offers support and training to country representatives to help them publish more data and enhance data quality.
- Second, to delve into the specific barriers faced by re-users, the results of the EDP Re-using Open Data report[8] were analysed. This report was based on an online survey, 33 in-depth interviews with companies using Open Data, and a short survey conducted at the International Open Data Conference (IODC) in October 2016.
- Third, to add to the insights gathered from both data publishers and data re-users, a broader scoped desk research was conducted. To complement the limited academic literature available on this topic, research also focused on sentiment analysis and discussions conducted with the Open Data community during conference and workshops.

Chapter 2 presents a deep dive into the barriers encountered by both data publishers and users of Open Data. Possible ways to overcome some of these barriers are described in Chapter 3. Finally, Chapter 4 provides a conclusion.

---

[3] European Data Portal, Open Data Maturity in Europe 2016

[4] Open Data Barometer 2015; Open Data Barometer 2016

[5] European Data Portal, Open Data Maturity in Europe 2016

[6] European Data Portal, 2017, Re-using Open Data

[7] European Data Portal, Open Data Maturity in Europe 2016; European Data Portal, Open Data Maturity in Europe 2015; European Data Portal, 2017, Recommendations for Open Data Portals

[8] European Data Portal, 2017, Re-using Open Data

2   Deep dive into the barriers faced by Open Data suppliers and users

Open Data offers a substantial economic and societal potential. However, this can only be realised when the data is accessible and re-usable in order to actually be re-used by organisations transforming it into value. In order to exploit this potential, it is important to gain insights on the one hand into the ability of organisations to seamlessly use the available data, and on the other hand the ability of data holders to provide their data as Open Data. As stated in various reports[9], still many challenges persist both on the side of the data supplier and the data user to reap the full potential of Open Data. But what kind of challenges are these? And why do they persist? This chapter explores these questions by taking a deep dive into the particular challenges which both data suppliers and data users face in their work with Open Data.

2.1   Overview of the barriers

Although more and more countries are embracing Open Data, there are still many persisting challenges when developing Open Data initiatives. During a recent research[10], the EU28+ countries had to indicate which barriers are applicable to their national Open Data portal team when publishing data. They also had to indicate which barriers prevent potential re-users from re-using Open Data. Both results are shown in Figure 2.
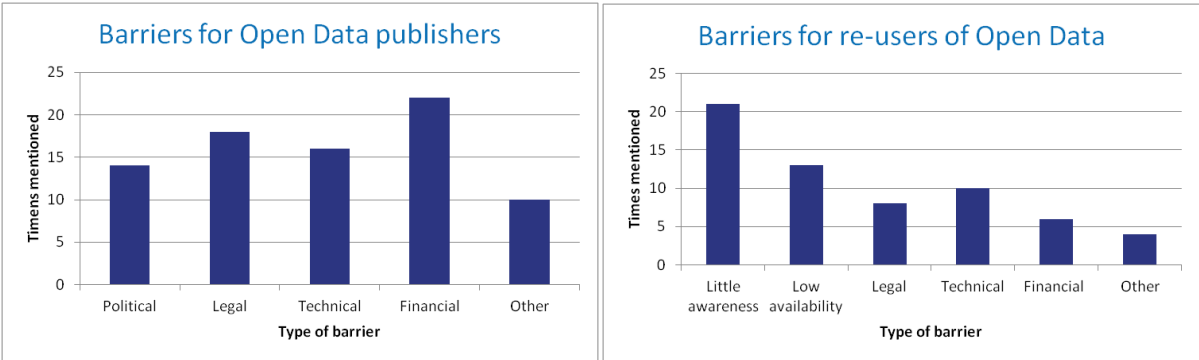
Figure 2: Barriers for Open Data publishers and barriers for re-users of Open Data

Regarding the barriers that hamper the publishing of Open Data, Figure 2 shows that the financial barrier is the most important barrier, mentioned by 71% of the countries. This barrier refers to the funding needed to deploy Open Data activities such as managing an Open Data Portal. But also political, legal and technical barriers play an important role. They are all mentioned by at least 45% of 31 European countries surveyed. When zooming in on the barriers faced by re-users according to national Open Data teams, a clear conclusion can be drawn. A lack of awareness clearly hinders the user community from reaping the full benefits of Open Data. Contrary to data providers, data re-users see the financial barrier as being one of the least important barriers. Followed by legal barriers which appears as second most important barrier for publishers. In addition, the low availability of particular data, indicated by 42% of the respondents, also shows room for improvement on the side of the data supplier.

Based on the bilateral interactions with Member States and desk research, another barrier was identified, namely the organisational barrier. This refers to the process of institutionalising Open

---

[9] European Data Portal, Open Data Maturity in Europe 2016; European Data Portal, 2017, Re-using Open Data

[10] European Data Portal, Open Data Maturity in Europe 2016

Data within organisations. Figure 3 summarises the most occurring barriers that prevent both data publishers and data re-users from efficiently releasing or re-using Open Data. The low availability of Open Data is covered under 'technical barriers' in this report.
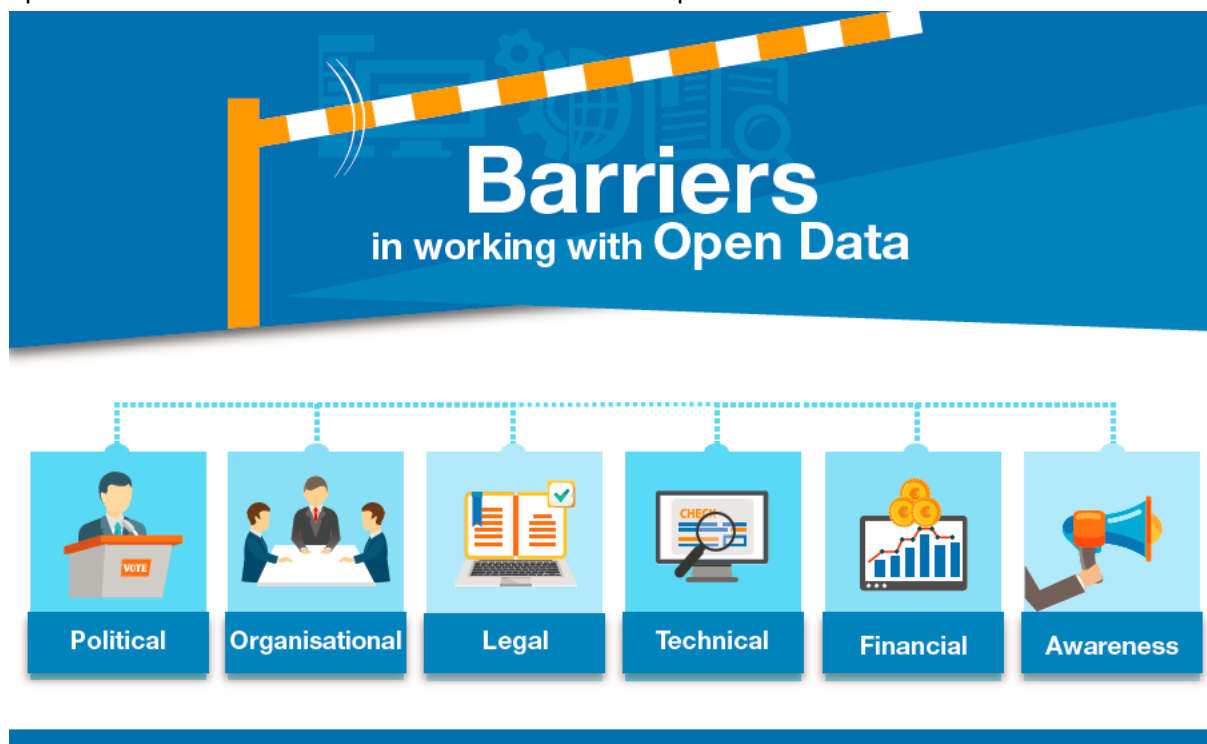

Figure 3: Barriers encountered when working with Open Data

As shown in Figure 3, the barriers faced can be clustered into six non-hierarchical categories:

1. **Political**: Politicians and policy makers determine the priority of Open Data by incorporating it in their plans but this priority has to descend into the administrations where it should be implemented. Likewise, in the private sector, managers may be reluctant to invest resources in exploring the potential value of Open Data because of its novelty or lack of proven monetary benefits.

2. **Organisational**: The organisational perspective examines the ways in which organisations enable or constrain the publication and use of Open Data. With regards to the internal structure of organisations, releasing Open Data should become an integral part of the operational processes. In this process, all actors involved need to be organised and well aware of their responsibilities. However, the organisational perspective also includes external elements, such as the cooperation and interaction between the various Open Data stakeholders.

3. **Financial**: Financial resources are needed in order to release and publish Open Data. For some government departments, such as many Chamber of Commerce's in Europe, moving towards Open Data would even mean losing a source of revenue as they would have to stop charging for data that it makes available. On the user side, financial considerations might also play a role in determining whether potential re-users decide to make use of certain Open Data as it will need a positive business case. Although Open Data is available for free, using it and packaging it into an application may require an investment.

4. **Awareness**: The value and potential benefits of Open Data for companies are not always recognised or known by data publishers or potential users. A low awareness around the use

of Open Data leads to a publication strategy not suited to the needs of users and potential data users not aware of the availability of certain Open Data sets. Raising awareness is therefore crucial in reaping the full benefits of the Open Data potential.

5. **Legal**: A legal framework provides a basis for Open Data, for example when it makes opening up data mandatory. In the European Union, the framework can be constructed based on the PSI Directive[11], but the legal basis might not be everywhere as specific as in the European Union. Working with Open Data requires that users have to know whether and under which conditions they can use the information. It should therefore be clear which licence applies to a dataset and what it permits.

6. **Technical**: The technical perspective focuses on the availability of Open Data, the infrastructure and technologies to either publish or make use of Open Data. It also brings into focus the requirements on the user side for being able to retrieve the data, processing it and making use of Open Data. Working with Open Data, both on the supplier and the user side, requires particular expertise and skills in the area of IT and data processing as well as data management.

2.2 Political barriers

Political barriers emphasise the importance of political developments and political differences between countries. Approximately one-third of the EU28+ are facing political barriers[12] with regards to data publishing. Three political barriers can be identified on the side of the data publisher and one on the side of the user of the data.

A first political barrier relates to the **lack of political will**. Open Data, although being included in the Public Sector Information Directive (see section 2.4), is not always a priority on the agenda of a policy maker. This might be caused by

33% One-third of the EU28+
Is facing **political** barriers

Figure 4: Number of EU28+ countries facing political barriers

unfamiliarity with the concept of Open Data, or not seeing the value and potential benefits of the reuse of Open Data. Among data publishers, this may result in a lack of resources available for moving forward with Open Data. When a government puts Open Data on the agenda, resources and funding for Open Data will most likely be made available as a result. In addition, political willingness and support are key in making otherwise reluctant administrations publish their data.

"One of the main barriers to Open Data is the lack of awareness of the benefits of Open Data among politicians. A political decision is needed to grant government funding to the institutions that are now partially financed from selling data."

A second political barrier relates to the **lack of awareness by both politicians and lower levels of government** or public administrations. Open Data portals are often top-down imposed, as part of a new national Open Data initiative driven by the national (federal) government[13]. An important cultural challenge within public ad-
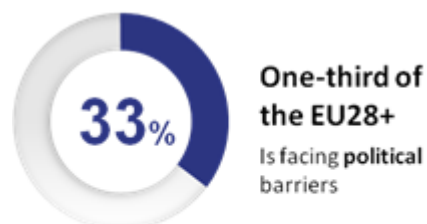
---

[11] EUR-Lex, 2013, Revision of PSI Directive
[12] European Data Portal, Open Data Maturity in Europe 2016
[13] European Data Portal, 2017, Recommendations for Open Data Portals

ministrations is the fear to lose control over data[14]. The public sector needs to be convinced of the benefits first by making civil servants aware of the rationale of sharing. Managing cultural change is therefore an important aspect of the implementation of an Open Data policy. Open Data should not be seen as an extra activity, but rather as an integral part of the day-to-day activities. Both government agencies' staff members and political leaders need to shift their mind-set about data from one that has Open Data as an exception to one where Open Data is the default. Politicians are generally not well aware of the benefits of Open Data, resulting in the fact that they do not identify Open Data as a priority, but solely see it as a 'nice to have' feature. Since Open Data is driven by public body representatives, it is essential to make political leaders more aware of the benefits Open Data brings to society.

Third, the **political structure of countries** can impact the efficient roll out of Open Data through the administrations. Countries with a federal political system have different regions with high autonomy and perhaps diverging policies. In countries such as Germany and Belgium this results in these regions having their own Open Data strategies and pertaining data portal. Keeping track of all these regional initiatives and coordination and cooperation between national and regional public sector bodies creates a challenge. Moreover, this can lead to a diverging use of standards, licences and formats, thereby creating additional technical or legal barriers. In countries with such a federal structure, it is therefore important that these regional initiatives are coordinated at a national level, and that regional portals are integrated in the national portal. Substantial differences exist between the countries in terms of whether there is integration into a national portal or not. Austria for instance has integrated all its regional portals in the national portal. However, at the European level, only 13% of the European national portals have successfully overcome this barrier. Belgium, due to its federal system, has four governments that each have an Open Data policy and regulation. Even though these policies and portals are similar, it still makes it difficult to access and re-use certain important data sets across the whole country. An example of several Open Data initiatives existing parallel comes from Spain. Spain has integrated these initiatives in their national portal (Figure 5).

"Greater engagement of politicians in promoting the Open Data paradigm at all levels of central, regional and municipal level is needed"

Figure 5: Organising multiple Open Data initiatives in Spain[15]

---

[14] European Data Portal, 2015, e-Skills and Open Data

[15] http://datos.gob.es/en

Political barriers not only appear on the side of the data publisher, they can also affect re-users of Open Data. The decision to make use of Open Data and to rely on it as a resource is often taken at managerial level within companies and organisations. Therefore, the decision-makers within the organisation need to be **convinced of the added value** of Open Data. Second, it is a political decision to communicate about the use of Open Data externally. Companies can communicate about Open Data in their commercial activities (e.g. branding), and the decision to do so is taken on a managerial level. When drafting the report on the re-use of Open Data, most large organisations were reluctant to talk about how they use Open Data, whereas their services and products are clearly building upon Public Sector Information. Communicating about Open Data leads to more awareness about the benefits of Open Data, in turn leading to an increase in data publishing and data quality.

2.3 **Organisational barriers**

Organisational barriers include factors related to the internal and external organisation of data publishers and data users that constrain the publication or adoption of Open Data. These factors appear on both the Open Data publisher and user side. This perspective also touches upon the skills needed to work with open Data and the need for interaction with other Open Data stakeholders when organising the release or re-use of Open Data.

2.3.1 Institutionalising Open Data in public bodies and companies

First, from the perspective of a data supplier, institutionalising Open Data might be a challenge. **Management structures of Open Data portals tend to be ad hoc**, being the result of national Open Data policies[16]. To institutionalise the publication of Open Data, it should become an integral part of the data creation process, rather than a separate or additional activity to the daily operational processes and routines. In order to become sustainable, the governance model for an Open Data portal has to become embedded in 'business as usual' government functions, and has to be able to continuously adapt to changing government priorities[17]. This barrier can therefore also be seen as an opportunity to improve data governance, update workflows and to automate them as far as possible. An example comes from the United Kingdom, which has integrated the data holders in the data provision process. Automated reminders are in place for datasets that should be published according to their specific time series and whose publications have slipped.

"The process of data publishing needs too much manual effort and is not part of usual process."

Other organisational challenges for data publishers relate to **negotiations and considerations** that have to take place amongst the managers when Open Data policies are developed. These managers may have different interests or even resent the concept, slowing down the implementation of organisational Open Data policies. For instance, by having their data published on a specific data portal instead of on their own government portal, may be perceived as a loss of value and visibility of the organisation. In other cases, this particularly holds ground when the organisation needs to share information which is perceived sensitive. Some countries are struggling with these barriers

"A better coordination is needed between public sector bodies."

---

[16] European Data Portal, 2017, Recommendations for Open Data Portals
[17] European Data Portal, 2017, Recommendations for Open Data Portals

as the roles and responsibilities of the different administrations and or agencies may not have been sufficiently clearly outlined in the initial Open Data Strategy. This can equally be the case regarding data domains where national debates have been seen to take place in a number of countries as to whether a certain domain/ministry would fall under the Open Data initiative.

Organisational challenges also apply to businesses that wish to make use of Open Data. In a recent report[18], aligning the internal organisation with the use of Open Data was mentioned as one of the success factors for using Open Data. Businesses need to integrate the use of Open Data into their workflows and organisation. This means that **the use of Open Data has to be integrated into existing operational processes**, or that new processes are started with Open Data. This implies ensuring that

> *"There is a lack of response and interest of administrative bodies in understanding the needs of infomediary businesses."*

data collection is operated either directly, via aggregators (see 2.3.3), or collected and then curated. Not only does this change require resources, both financially and in terms of employees, but it also requires a new organisational set-up. Roles and responsibilities have to be clearly defined, and in case of working with partners in the data value chain, cooperation needs to be initiated and managed. These elements could be challenging for companies. A key challenge for example is to recognise a business challenge and to build a team capable to address that challenge with Open Data[19].

Furthermore, public administrations have been seen over time as being the first re-users of their own Public Sector Information. The Once Only Principle of having data collected once and then shared and re-used across different services should become the default[20]. However, approximately half of the public services (49%) use authentic sources for pre-filling online services according to the eGovernment Benchmark report for 2016[21]. When for instance submitting a tax form online, this form can be partially pre-filled based on information from other organisations or previous tax returns. Here, sharing data across public administrations does not only stimulate efficiency in avoiding to collect data several times. Re-use of data is also a means to transform and rethink processes and public services delivery. However, efficiency gains remain limited as organisational barriers prevent a full restructuring of processes.

2.3.2   Skills needed for working with Open Data

A **lack of skills** among several government departments and lower government levels to deal with Open Data is a barrier. Public bodies should have the required technical knowledge to ensure a smooth and automated data release process, which is not always the case[22]. This barrier is more prominent in federal countries, where national authorities depend on regional and local governments for the supply of Open Data. On a central level (e.g. data management department or project team in charge of a national portal), one might have a correct understanding of Open Data and the appropriate skills to work with Open Data, but at the same time one is dependent on several governmental institutions and regional and local governments in the data provision process.

---

[18] European Data Portal, 2017, Re-using Open Data (company: BBVA)

[19] European Data Portal, 2017, Re-using Open Data (company: eGovlab)

[20] EU eGovernment Action Plan 2016-2020: Accelerating the digital transformation of government

[21] eGovernment Benchmark 2016 Insight Report

[22] European Data Portal, Open Data Maturity in Europe 2016

Users **working with Open Data require certain skills to fully make use of its potential**.[23]. These skills consist of both hard and soft skills which relate to technical and respectively management skills as shown by Figure 6. Working with Open Data requires analytical, statistical and technological skills, next to communication skills and domain knowledge. Many companies mentioned that the hard skills are needed to find the relevant Open Data, make this data usable and customise it, as the format and structure might differ from dataset to dataset. At the same time, companies mentioned a need of soft skills to recognise how Open Data can solve specific business challenges. Having an open mind was mentioned as being equally important, which refers to the ability to look at data and to identify the economic and societal added value of using this particular dataset. This plays in tune with an earlier report published by the EDP on the type of skills needed to work with Open Data in which the typology between hard and soft skills is further developed. [24]
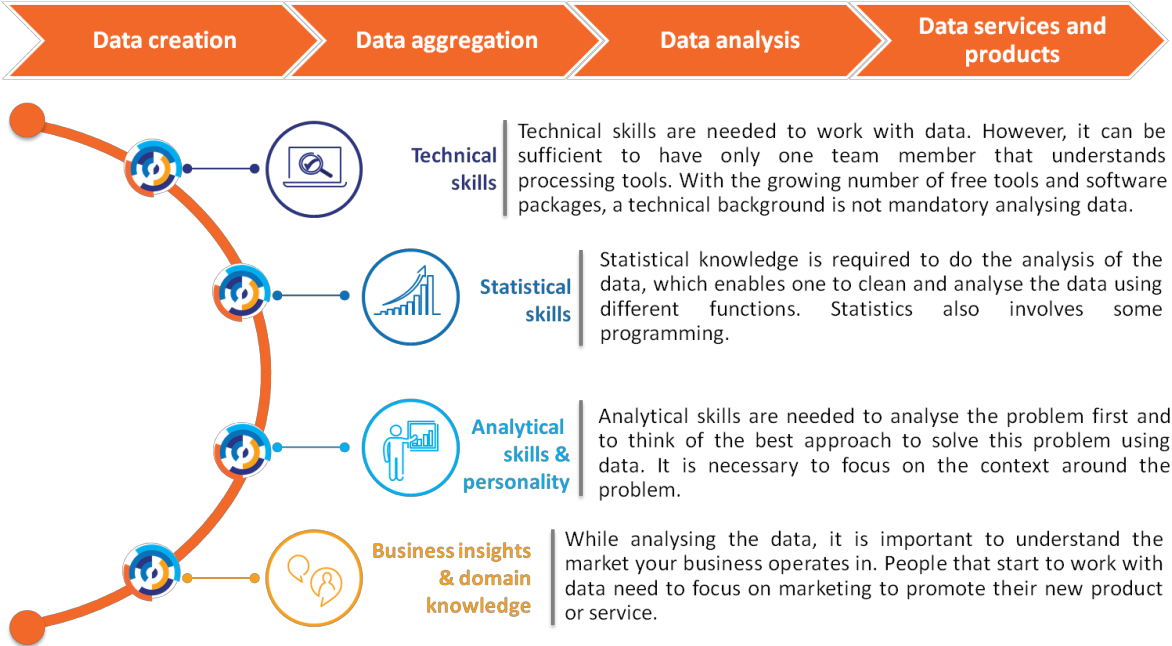

Figure 6: The skills needed to work with Open Data

These skills however, are not always readily available within public administrations. Especially lower levels of public administrations (municipalities, regions) might suffer from a lack of skills to efficiently produce and publish Open Data. The same accounts for re-users of Open Data, where the mentioned combination of both hard and soft skills is not always readily available. Although IT knowledge is often available, re-users indicate to miss marketing skills and business insights[25].

2.3.3   Interaction between departments and Open Data stakeholders

The organisational perspective also includes a social element, bringing into focus the importance of interaction between the various stakeholders in the Open Data Value Chain[26]. This value chain, as illustrated in Figure 7, displays the various steps by which raw data is transformed into (economic) value.

---

[23] European Data Portal, 2015, e-Skills and Open Data

[24] European Data Portal, 2015, E-skills and Open Data

[25] European Data Portal, 2017, Re-using Open Data

[26] MEPSIR (2006), p. 46 and European Commission, 2013, elements of a data value chain
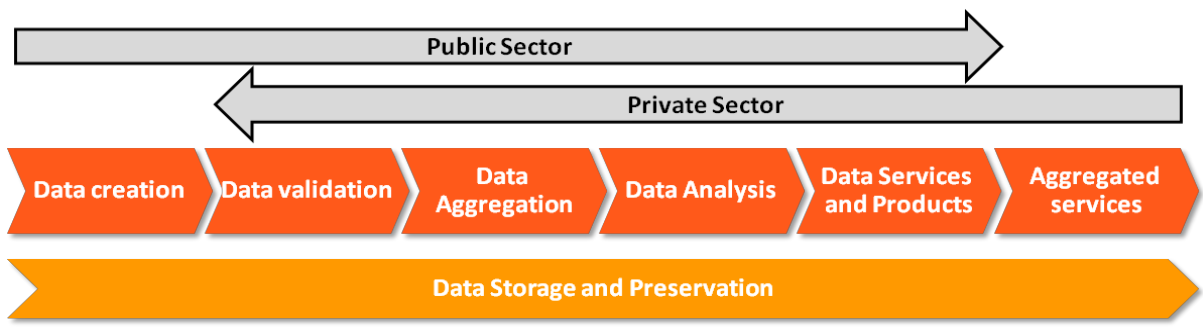
Figure 7: Open Data value chain (source: re-use report)

The first step in the value chain is to create data. The data is then validated and released, for example through a portal or bought by a private company, after which it can be analysed. By aggregating different data sets, new data is created which can lead to new data services or products. Finally, these data services and products can be further aggregated. This process is visualised below in Figure 8. Throughout this process various actors are involved. Following the steps above, data creation is done by the Suppliers. The data is subsequently collected and aggregated by the so-called Aggregators. Developers then use the data for the development of new applications, and Enrichers use data to gain new and better insights from the analysis of the data. Enablers facilitate the supply or use of Open Data for the other archetypes.
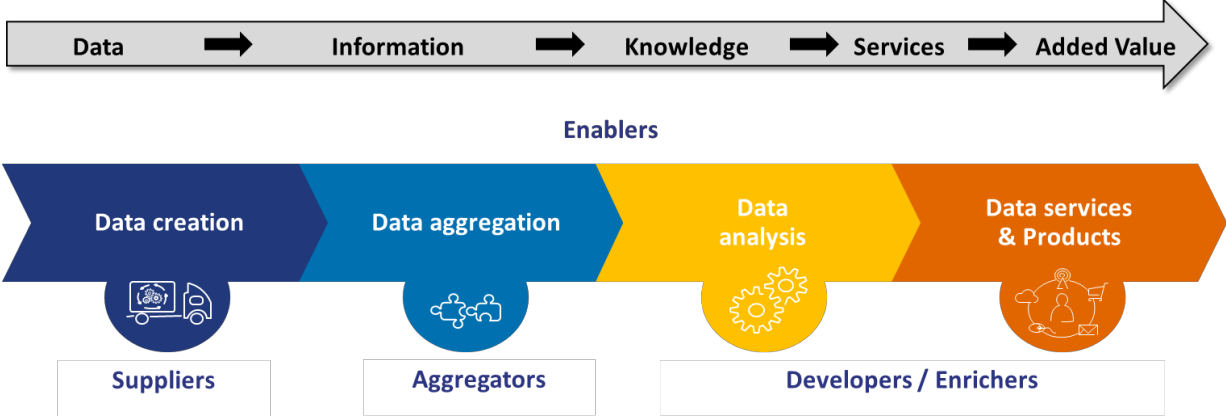

Figure 8: Data value chain archetypes (source: re-use report)

The Open Data value chain brings many interdependencies between the different actors in the chain. Managing these interdependencies requires **a certain level of coordination**, including interaction between the stakeholders. Data providers may require interactions with the potential users of Open Data to determine which data is most valuable and should be opened first. At the same time, data users may require to interact with data suppliers to inform them of issues such as insufficient metadata. For re-users that rely on systematic and continuously published data, it is important that they can trust the data supplier to publish it periodically. If this is not clearly described, the data user may require to interact with the data supplier in order to get this confirmation. Those types of interaction can be challenging, since Open Data suppliers, users and policy makers may not know each other, and Open Data infrastructures may not always allow for interaction between the stakeholders involved.

2.4 Legal barriers

The legal barrier focuses on challenges related to Open Data legislation, Open Data policies, open government directives and licences. In 2003, the European Commission adopted legislation to foster the re-use of Public Open Data in Member States via the Public Sector Information (PSI) Directive 2003/98/EC[27]. The main objective was to ensure equal treatment of all potential re-users where the public sector body had released information for re-use. A revision of the PSI Directive was introduced in 2013 (Directive 2013/37/EU[28]). The main amendments are the adoption of the "open by default" principle, the breakaway from cost-based charging for PSI towards a marginal cost-oriented fee and increased transparency regarding the calculation of fees, the inclusion of certain cultural institutions as public sector bodies, and support to machine-readable and open formats. Member States were obliged to transpose Directive 2013/37/EU by 18 July 2015. By the end of 2016, 96% of the EU Member States have transposed, partly or completely, the revised PSI Directive[29].

"There is no legal basis defined. The lack of a legal framework is restricting the re-use of Open Data."

First, it is important that there is a clear legal framework in each country, underpinning and supporting the release of data. From the perspective of the data provider, the legal framework for opening datasets may be **unclear, unspecific or not even developed**. The revised PSI Directive serves as a clear guide for EU countries. Although it is not mandatory for the EFTA countries to complete the transposition, Liechtenstein and Norway are planning to finalise the transposition by spring 2017. Again, here we notice that from a legal perspective, PSI and thereby Open Data is being recognised more and more, thereby paving the way for further publication of data. Especially federal countries, such as Belgium and Spain, indicate that the legal framework in their country is not always clear, with different regional governments each having an own Open Data policy and regulation.

A next challenge related to legislation are **privacy constraints** that prevent data publication. Multiple countries such as Belgium, Germany and Spain explained that a privacy framework prevents the publication of data, which is deemed privacy sensitive. Also in the Netherlands, one of the main legal barriers is related to privacy issues. Privacy is becoming an increasingly hot topic in the political and societal debate and can lead to restrictions when politicians and citizens do not understand that Open Data is not always privacy sensitive; an issue which becomes worse in the absence of an actionable legal framework as described above. Furthermore, the General Data Protection Regulation[30] may serve as a barrier for re-using data, when combining Open Data and personal data. Depending on the specific use and aim, this Regulation may restrict re-users from using private data if they do not fulfil the requirements of the Regulation.

"We have a strong privacy legal framework that makes the publication of some data sets difficult. Furthermore, we are a federal country, resulting in different governments with each an open data policy and regulation."

---

[27] EUR-Lex, 2003, Directive 2003/98/EC of the European Parliament and of the Council
[28] EUR-Lex, 2014, Directive 2013/37/EU of the European Parliament and of the Council
[29] European Data portal, 2015, Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources
[30] EUR-Lex, 2016, Regulation (EU) 2016/679 of the Parliament and of the Council

Third, one of the legal barriers that existed in Ireland was related to ensuring that the **appropriate licence** is applied. Without an open licence, data cannot be used freely. In order to overcome this challenge, Ireland has organised a public consultation on the appropriate licence for the Open Data initiative, leading to the selection of CC-BY. In some countries licences are not regulated nationally. For example in Germany, although there are national licences, those licences are not always used by regional governments or other public bodies.

From the perspective of data users, it may be unclear which licence applies to a dataset. A variety of licences exists. Without a licence, data is not truly open. A licence informs potential users that they can access, use and share the respective data. It provides users with certainty that the data can be used and shared for a wide range of purposes. Without a licence, data might be 'publicly available', but users will not have permission to access, use and share it under copyright or database laws. If it is unclear under which conditions data can be re-used, potential users are reluctant to use this data. It may not be clear which licence applies if various datasets with multiple licences are combined. When using several different datasets, it is important to check the compatibility of the licences assigned to these sets. In addition to global licensing standards, many countries have developed their own national licences. Germany for instance developed its own licence, based on the CC-BY licence. Switzerland does not use a licence, but has developed specific terms of use.

2.5 Technical barriers

The technical perspective focuses on the importance of Open Data technologies, platforms and infrastructures. Various technical barriers prevent both data suppliers and data users from efficiently working with Open Data. At the IODC 2016, re-users of Open Data were asked to indicate which technical barriers they face. The **quality and the availability of Open Data** were mentioned as being the most important barriers. Although a lot of data has been made available in the past years, resulting in over half a million data sets referenced at the EDP[31], the availability of Open Data persists to be a barrier. The quality of Open Data appears to be even more problematic, with data being published in different structures and in different formats.
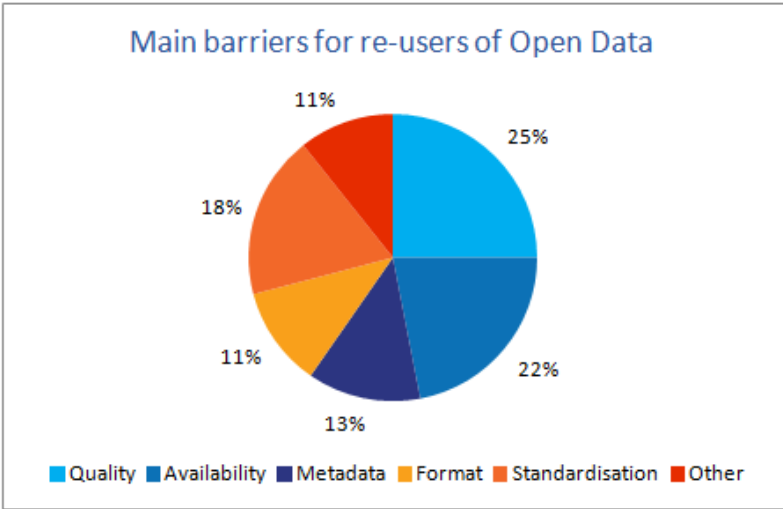

Figure 9: Main barriers for re-users of Open Data[32]

---

[31] https://www.europeandataportal.eu/data/en/dataset
[32] European Data Portal, 2017, Re-using Open Data

This chapter zooms in on these technical aspects. Chapter 2.5.2 describes barriers related to the quality of Open Data, and chapter 2.5.1 describes barriers related to the availability of Open Data. Barriers related to the metadata, the format in which Open Data is published and standardisation (as shown in Figure 9) are included in these two sections. But first, this section explores a specific data domain within the Open Data ecosystem in which technical barriers are particularly present: the geo-spatial domain.

A geospatial perspective on barriers

A significant part of all information used and published by public administrations and ex-changed with the public refers to specific locations. Its quality depends on the availability of 'spatial data', which is collected and linked (geo-referenced) to location. This is known as geo-spatial data. Geospatial data is among the five thematic data domains that are expected to have the highest economic impact, as they are expected to represent those with the highest demand from re-users across the EU[33]. However, there are a number of barriers that prevent users and publishers from efficiently working with geospatial data.

Different standards stemming from the INSPIRE Directive
The European geospatial community is driven by the INSPIRE Directive[34], which came into force in 2007. This Directive aims to establish an infrastructure for spatial information in Europe that is geared to help to make spatial or geographical information more accessible and interoperable for a wide range of purposes supporting sustainable development. As a result, the geospatial community has standards that differ from the standards applicable in the Open Data community. Data access is facilitated through standardised service interfaces for view-ing, downloading and transforming data[35] in the geospatial world, and data discovery is facili-tated through standardised metadata[36] and discovery services. Every geospatial dataset has a universally unique identifier (UUID), which is not always properly reflected in Open Data por-tals. Also with regards to formats, differences exist. The standard for publishing data in the geospatial community is a view service, with not necessarily having the raw data available for download. Within Open Data portals, usually only this view service appears because the IN-SPIRE Directive does not require the raw data to be available.

High entry barrier
The entry barrier for working with geospatial data is high[37]. One has to have a thorough un-derstanding of the several steps in geospatial data management and the formats that exist, as sourcing and translating complex geospatial datasets from a variety of sources and servers can be challenging. As the formats differ from those applicable in the 'normal' Open Data community, a lack of technical knowledge hampers Open Data specialists to work with geo-spatial data.

---

[33] European Commission, 2014, Guidelines on recommended standard licences, datasets and charging for the reuse of documents

[34] European Commission, 2017, INSPIRE

[35] Commission Regulation (EC) No 976/2009 as regards INSPIRE Network Services, OJ L 274, 20.10.2009, p. 9

[36] Commission Regulation (EC) No 1205/2008 as regards INSPIRE metadata, OJ L 326, 4.12.2008, p. 12

[37] Source: interview con terra, February 2017

Geospatial data not classified as geospatial data

There is already a lot of geospatial data available, but it is not often recognised and labelled as geospatial data[38]. Geospatial is sometimes classified differently by institutions. Think of data about air pollution that might be classified as weather data. Or data on the location of businesses that is classified as business information. In these cases, different standards and references are applicable, missing for instance a standardised spatial reference. One is not adhering to the excessive standards of the INSPIRE Directive, and data is published in different formats. This makes it difficult when in the end this data is needed to produce maps. This way, one is not profiting from the harmonisation efforts as articulated in the INSPIRE Directive.

Geospatial data is often still charged for

Geospatial data used to be (and still is) a pre-eminent source of income for national and regional governments. In Germany for instance, only two out of the sixteen States have opened up their geospatial data[39]. Having data available under an open licence is not required according to the INSPIRE Directive.

File sizes

Geospatial datasets might suffer from enormous file sizes, especially when considering satellite data. Having terabytes of geospatial data might pose a challenge for the storage and discovery of Earth Observation data, as Open Data portals need to have the capacity to store the data or deal with specialised metadata profiles for Earth Observation data. It also restricts the re-use of these datasets as specialised (and potentially expensive) download tools are needed to extract and process the files.

2.5.1 Quality of Open Data

The quality of Open Data and accompanying metadata is important. High quality and consistent (meta) data saves the re-users time as it facilitates automated processes and creates trust among re-users that they can rely on the data. A recent study[40], however, found that data users still perceive the **quality of Open Data to be low**. This is confirmed in another recent study on the sustainability of Open Data portals, in which portal owners in Italy, Romania and Spain indicated that data-quality issues continue to be a barrier to the use of data accessed via their portal, and that no hard levers exist to enforce a better quality[41]. The low quality refers to both the data itself as well as the accompanying metadata. An Open Data portal should not be a source of broken links, out of date and unused data nor poor metadata. Examples were mentioned where the indicators within a dataset or metadata describing the dataset did not correctly represent the actual data. This makes it difficult for users to find the data they are looking for as their queries do not find the sets which hold this infor-

"Data is often available in PDF only or it is difficult to obtain data, hence in non-machine readable formats. This makes the data less useful for re-users."

---

[38] Source: interview con terra, February 2017

[39] Source: interview con terra, February 2017.

[40] European Data Portal, 2017, Re-using Open Data

[41] European Data Portal, 2017, Recommendations for Open Data Portals
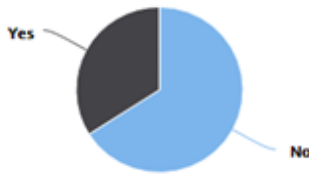
Ratio machine
readable datasets

Figure 10: Ratio machine-readable data sets on the European Data Portal

mation. As an explanation for the low quality of Open Data, it was mentioned that the provision of Open Data is still not a priority among data providers, creating limited stimulus for these bodies to publish consistent and high quality data.

The plurality of Open Data, being available in different formats, with different licences in different languages is another restraining factor brought forward by companies re-using Open Data. This stark heterogeneity of the various characteristics of Open Data restricts the usability of Open Data. And a **lack of standardisation** restricts the opportunity for users to develop permanent solutions to re-use Open Data in their processes.

The **interoperability of Open Data portals** is a challenge for portal owners and a barrier for re-users of Open Data. This refers to different portals using the same standards, such as the metadata standard DCAT-AP[42], which facilitates the sharing of data between different data catalogues. The EDP uses the DCAT-AP, but not all portals harvested by the EDP map their datasets in the same categories. A different mapping limits finding datasets on similar subjects between portals. Using common Open Data, metadata standards and application profiles will help to maximise the discoverability of data across portals.

Not only data publishers struggle with technical aspects. The **machine-readability** of datasets is a barrier for companies re-using Open Data. Machine-readable data refers to data that can be automatically read and processed by a computer. Examples of formats that are machine-readable are CSV, XLM or XLS. PDF files, on the contrary, are not machine-readable. Machine-readable formats can be accessed in an automated fashion. However, in 2016, only 52% of the EU28+ countries indicated to have more than 90% of their datasets in a machine-readable format according to the Open Data Maturity in Europe report 2016[43]. Moreover, according to the Metadata Quality Dashboard of the EDP[44], currently only 34% of the 620,000 datasets available on this portal are machine-readable. A dataset is considered as machine-readable if at least one of its distributions uses a machine-readable format. This low amount of machine-readable datasets is a problem, because it hampers automated processing of the data.

> *"Currently, the metadata interoperability for Open Government Data is high, but the interoperability of the data sources themselves is an important issue. In some areas these need to be solved before a widespread use is possible."*

A relatively low amount of machine-readable datasets might be caused by a focus on quantity by the publication team, rather than on quality. Teams may focus too much on publishing the 'low hanging fruit' – old or out-of-date data, aggregated data, or data that is not heavily used within the organisation. This is a problem, because if the published data is not relevant to potential users, it is not being used, and benefits are not seen by those involved in the initiative, resulting in cynicism[45]. Figure 11

---

42 World Wide Web Consortium, DCAT W3C Recommendation 16 January 2014, https://www.w3.org/TR/vocab-dcat/
43 European Data Portal, Open Data Maturity in Europe 2016
44 European Data Portal, MQA service. The metadata quality is checked on a weekly basis. This data is based on the check of 1 March 2017
45 European Data Portal, 2017, Recommendations for Open Data Portals

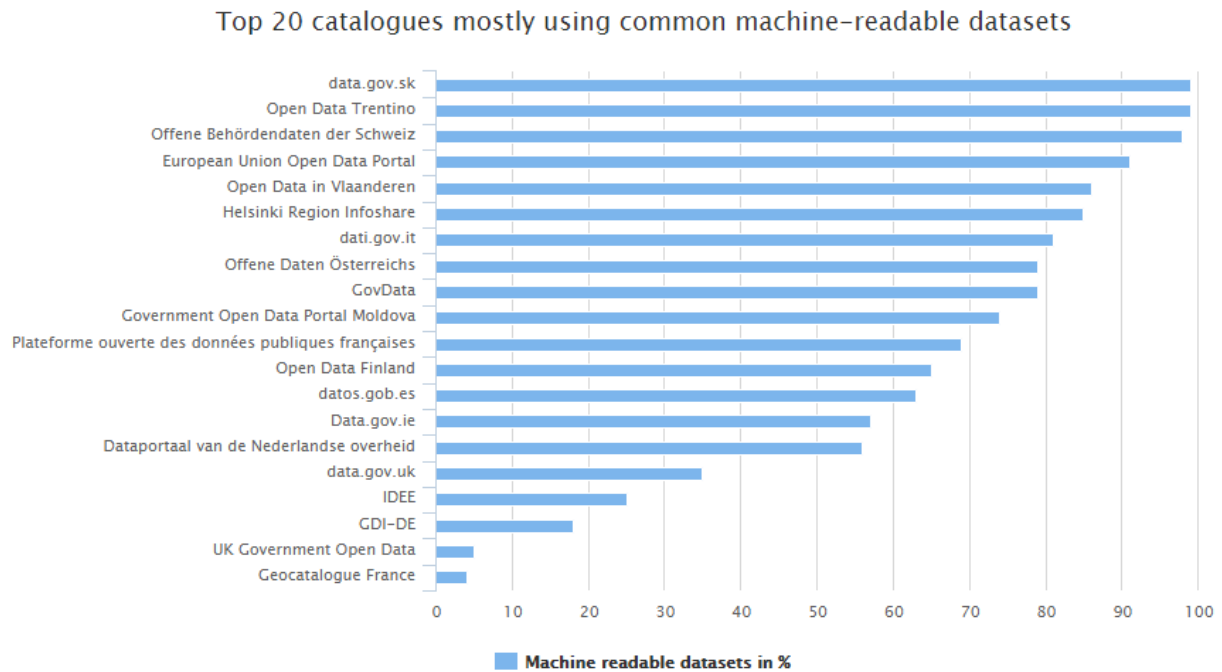shows the percentage of machine-readable datasets for the top 20 catalogues mostly using machine-readable datasets.

Figure 11: Top 20 catalogues mostly using common machine-readable datasets[46]

Another barrier for companies re-using Open Data has to do with the **quality of the metadata**. Some companies, such as CropDiagnosis[47], rely on the systematic and continued publication of Open Data. It should be clearly reflected in the metadata whether Open Data is periodically updated, including its frequency. Unclear periodical updates lead to doubts about the usability of a dataset. Companies re-using Open Data need to be able to trust data suppliers, as they base their business model on the use of particular datasets. On data.gov.uk, only 25% of the datasets includes information about its update frequency[48]. When the release frequency is unclear, this could prevent companies from re-using this data. Furthermore, different metadata structures are used among different catalogues, limiting the automated processing of (new) data sets. Data structures may also vary over time. Companies using data from such catalogues indicate to spend more time on quality assessments before further using the found data[49].

2.5.2 Availability of Open Data

Countries are maturing on their Open Data journey[50]. Open Data policies are being developed and data sets from multiple government departments are being opened, resulting in more and more data sets being made available. However, despite these encouraging developments, evidence shows that a lot of work remains to be done in this area, as the availability of Open Data remains to be a chal-

---

46 European Data Portal, MQA service

47 CropDiagnosis

48 European Data Portal, 2017, Recommendations for Open Data Portals

49 European Data Portal, 2017, Re-using Open Data

50 European Data Portal, Open Data Maturity in Europe 2016

lenge, at least in the perception of re-users. 73% of the companies using Open Data indicate to find it difficult or very difficult to find the data they need[51] (Figure 12).

Is it easy to find the data you need?

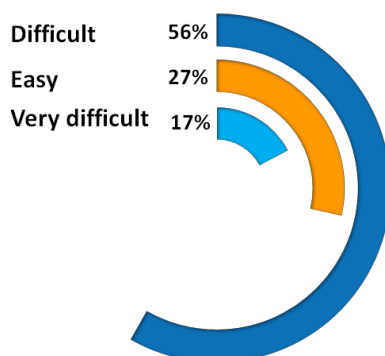|	
Difficult	**56%**
Easy	**27%**
Very difficult	**17%**

Figure 12: The ease of finding data by companies

The **poor discoverability of the data** users are looking for is a barrier which is also related to the low levels of quality in the descriptions of the data sets themselves and the plurality of platforms where the data can be found. When the descriptions are not specific enough or when the dataset is disseminated on a particular platform, it is hard for the user to find the data they need even when it is published. The dispersion of data, among several platforms especially, hinders the discoverability of data, when there is a lack of standardisation in metadata. When dispersed among platforms in different countries, also language barriers might play a role in the poor discoverability of Open Data. This is even more prominent when portals have only limited or very basic search functionalities.

2.6   **Financial barriers**

It always comes down to the money. For Open Data, the financial benefits are substantial. The direct market size of Open Data is expected to be 59.7 bn EUR for the EU28+ in 2017. Companies can therefore turn Open Data into economic value by either the enhancement of their operational processes or the development of new services or products. For public administrations, Open Data results in significant cost savings. In total, the cost savings for the EU28+ in 2020 are forecasted to be 1.7 billion EUR[52]. At the same time, arguments related to financial issues still hinder Open Data to move forward. These financial barriers are most prominent at the side of data publishers.

"It is a cultural issue that Open Data is not yet considered as an asset by all

Public authorities can be used to selling particular types of data, such as cadastral data. The UK Office for National Statistics for instance charges for particular data services to major clients, such as the Central Bank. When legislation forces them to open up their data free of charge (or at marginal costs), these authorities face a loss of income. In these cases, it is therefore necessary to reorganise the funding model of the public body. The **benefits of publishing Open Data for free are not always clearly documented**, making it difficult for administrations to justify the loss of revenue, or more broadly, understand the benefit of publishing data in the first place. In these cases, public administrations might not be inclined to open up their data.

---

[51] European Data Portal, 2017, Re-using Open Data
[52] European Data portal, 2015, Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources

96% of the EU

Member States have implemented the marginal (zero) cost model

Figure 13: The use of a marginal cost model in the EU

Public sector bodies in various countries have witnessed a strong growth in demand for information they provide after switching from cost-based pricing of Open Data to free or maximum marginal cost priced information (Figure 13). Cost-based pricing models do not bring cost savings to public authorities in the long run while free or marginal cost models are more beneficial. In most cases, the cost recovery model even creates barriers to the access and the re-use of Open Data[53]. Especially economically less strong people (e.g. citizens, students, start-ups etc) will not make use of Open Data if a fee is requested to obtain the data[54].

At the same time, **additional funding** might be needed for the implementation of an Open Data policy, which requires investments in infrastructure, people and processes. This implies that the implementation of the new strategy needs to be accompanied by training and awareness-raising activities. Although a lack of funding is still seen as one of the major barriers, in 2016, several countries indicated this barrier was being addressed by organising more meetings and negotiations to raise more awareness around the importance of Open Data, such as the IODC. [55]. In Estonia, organising events and raising awareness around both benefits and challenges has already led to the allocation of more resources for opening up data and pursuing communication efforts. This particular barrier links to the political barriers mentioned in section 2.2, as government funding depends on political decisions. Especially portal owners struggling with limited re-use of data published via the portal noted that it made funding and support for the portal difficult to maintain[56]. For regional portals, with less secure funding streams, this problem was more acute.

"Authorities, that are partly financed by the income of the sale of data, have to find other means to finance when publishing data free of charge. Another financial barrier is that the cost of publishing data can be hard to justify."

"The costs for the renewal of the old technical systems and their interfaces are barriers."

Sustainable funding could be a barrier for data providers, since the funding of a portal varies during the lifecycle of the portal. Sustainable funding is important to support day-to-day operations, allows future planning and acts as a mechanism to provide confidence to data users. Funding might become an issue when the momentum of launching an Open Data portal has been lost. While at the outset of an Open Data initiative, the Open Data portal might gain significant attention and support, maintaining funding for the portal can be difficult. Governments change, their focus goes elsewhere and priorities might shift. There are portal owners who describe their portals as being 'side projects', not being a priority and therefore running the risk of losing funding when government priorities shift. Multiple portal owners reported that their funding only supported basic day-to-day portal operations rather than new developments[57]. But

---

[53] Trojette (2013)

[54] European Data portal, 2015, Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources

[55] International Open Data Conference

[56] European Data Portal, 2017, Recommendations for Open Data Portals

[57] European Data Portal, 2017, Recommendations for Open Data Portals

EUROPEAN DATA PORTAL

not only costs related to launching and maintaining the portal from a technical perspective play a role. Ensuring data governance is another crucial element that does not come without costs. Data governance is key for ensuring a continuous flow of data to the Open Data portal, but it requires resources and capacity. The set-up of Data Portals is rather a novel task and sustaining these over time is a topic that has not yet been included in most, if not all countries.

Three different funding models can be distinguished. Most national portals are funded by a specific government department, such as a digital agency in Norway. Portals with this funding model have to 'compete' for funding with other digital projects, such as eGovernment portals offering access to online public services. This could be a barrier for the further development of the data portal. The German and Austrian national portals have federated funding agreements with several regions contributing to the national portal. Regional portals typically have mixed funding: The Spanish Aragon portal for instance is half-funded by the regional government and half-funded by the European Union. However, the latter funding stream is a temporary solution. When desired outcomes and conditions are attached to the funding, this could be another restraining factor for data publishers. At one particular national portal, it occurred that the large majority of their budget could only be spent on technical aspects of the portal, limiting their ability to run user-engagement activities, which they regarded as being more important for the future uptake of the data made available on the portal[58].

2.7 Awareness barriers

Open data has the potential to unleash innovation and transform every sector of the economy. But this can only be realised when both data publishers and data users are convinced of the benefits, resulting in more publicly available datasets and more Open Data being re-used. However, public awareness of Open Data and its relevance is still low, although this is slightly improving[59]. A reason for the low awareness is that Open Data is an abstract issue with unclear benefits to everyday life. The value drivers for opening up and sharing data are not necessarily clear to the public, and views on the benefits of Open Data are not universally shared within organisations and between potential re-users. Lastly, concerns may arise that focus on issues of privacy, anonymisation and consent.

On the supplier side, institutionalising Open Data might be a challenge, as indicated in section 2.3. It requires a cultural change within public administrations. Not all government departments and governmental levels (e.g. regions, municipalities) have the same understanding of the value drivers of Open Data. Imposing an Open Data policy without appropriate training and knowledge-sharing might result in Open Data being published in low quality, if made available at all. In addition, a lack of knowledge on the specific needs of re-users further hampers the re-use of Open Data by the private sector. Understanding the business needs for your portal, accompanied by a clear business plan, increases the potential for re-use and breaks the vicious circle of Open Data being published while not re-used and thus makes the portal more sustainable.

"The demand-side is critical for the development of Open Data."

---

[58] European Data Portal, 2017, Recommendations for Open Data Portals

[59] European Data portal, 2015, Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources

Figure 14: The lack of awareness among re-users, as indicated by EU28+ portal owners

According to 68% of the portal owners in the EU28+, data users also lack awareness of the available datasets and the potential benefits Open Data could bring. Interviewees of the 'Re-using Open Data' report[60] believe that there is already a goldmine of data available, waiting to be exploited. A better understanding of the potential value in the private sector could create a virtuous circle in which an increase in the use of Open Data stimulates the provision of more data. Re-users of Open Data also perceive their clients to have an either poor or mixed understanding of the concept of Open Data[61]. The function and position of the client's main contact person has a big impact on the understanding of Open Data. This ranges from a fear of sharing 'their' data to thinking the whole concept of sharing information without compensation is something for 'hippies'. Both of these findings indicate that there is still a lot of unfamiliarity with the concept of Open Data.

There are still only few awareness raising activities around the availability of Open Data[62]. Most citizens and businesses are not aware of existing publicly available data sets and the benefits of using Open Data. Entrepreneurs or other re-users of Open Data are not aware of the data that data holders actually have. By knowing how 'their' data can be used, data publishers could align their publication strategies with this need, resulting in the provision of better and more suitable Open Data for those organisations that can transform it into value. This underlines the importance of a continuous dialogue between data users and data holders, to stimulate both the publication and the re-use of data. More specifically, awareness needs to be raised regarding the availability of datasets, the specific needs of Open Data re-users and the value drivers of Open Data, hence the benefits of Open Data.

> *"Entrepreneurs do not know what data data-holders have. On the other hand, data-holders do not know what data entrepreneurs want."*

---

[60] European Data Portal, 2017, Re-using Open Data
[61] European Data Portal, 2017, Re-using Open Data
[62] European Data Portal, Open Data Maturity in Europe 2016

3 Best practices to overcome the barriers

The previous chapter showed that different types of barriers can be distinguished that prevent the Open Data community from reaping the full potential of Open Data. These barriers can be both present on the side of the data publisher or on the side of businesses, NGO's and public bodies re-using Open Data. This chapter describes best practices applied by either data publishers or re-users of Open Data to tackle some, but not all, of the barriers described in Chapter 2.

3.1 Best practices to overcome political barriers

Coordination and cooperation between national and regional governments is a key challenge for data publishers. The extent to which this barrier plays a role is strongly dependent on the structure of countries and the coordination mechanisms in place. Norway for example indicated that their regions are highly independent and are therefore not included in the national programme regarding Open Data. A similar situation occurs in Germany, because of the federal structure of the country.

Austria and Italy, countries with a federal structure, have developed solid solutions and coordination mechanisms to overcome barriers related to the collaboration with regional portals. Integrating the regional portals into the national portals increases the accessibility and visibility of Open Data throughout the country. The Austrian example shows how regional portals can be integrated in the national portal, with the regions being co-owners of the national portal and therefore sharing the responsibility for the national portal[63]. The Italian province of Trentino has also managed to integrate different kind of regional actors, such as provincial public administrations, municipalities and other agencies, in their data provision process (Figure 15)[64].
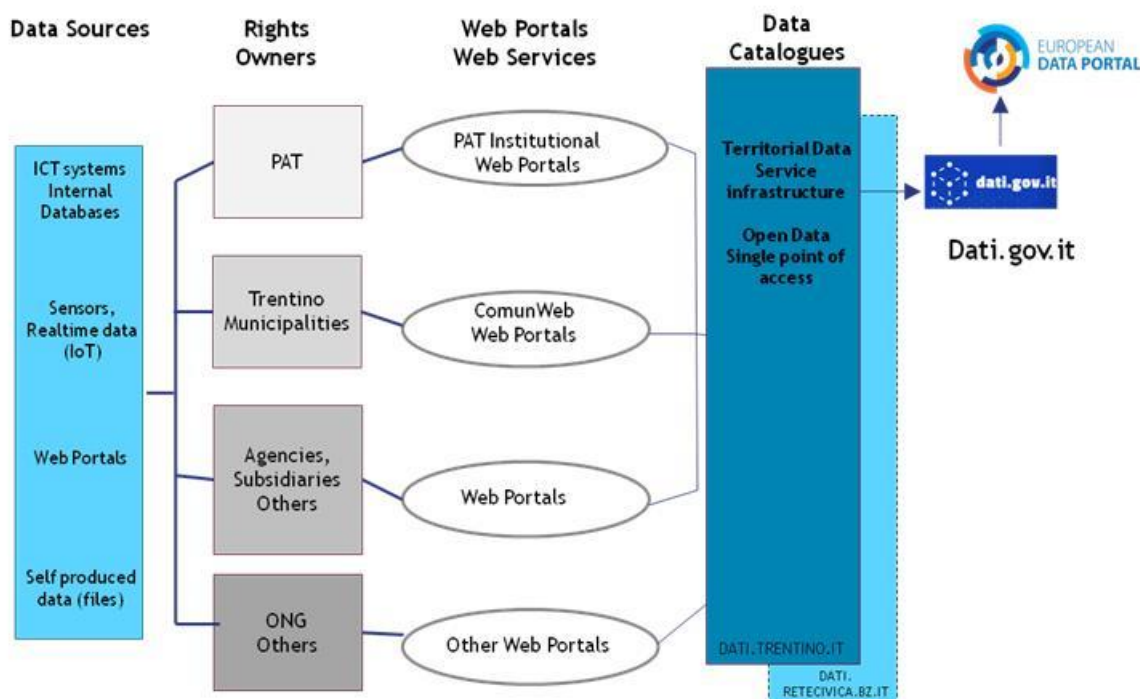

Figure 15: Sources of the Trentino Data Portal

---

[63] European Data Portal, Open Data Maturity in Europe 2016

[64] Source: Presentation 'Beyond the barriers to foster Digital Single Market', European Week of Cities and Regions, 12 October 2016

An example of how different actors within a national Open Data ecosystem can work together is the Six City Strategy from Finland.

> **Regional collaboration and national coordination in an Open Data ecosystem**
>
> In Finland, the six largest cities have joined their forces to solve urban challenges in the Six City Strategy (6Aika)[65]. Within the context of this framework, the cities of Helsinki, Espoo, Vantaa, Tampere, Turku and Oulu open up their data while developing and following shared operating models. These cities use standardised and open API's and shared licenses, and there is a strong focus on the re-use of Open Data by businesses. To ease the discovery and usage of data, three of these cities (Helsinki, Espoo and Vantaa) use a shared Open Data platform to publish their data: the Helsinki Region Infoshare platform[66]. The other cities each have an own Open Data portal, and all of these portals feed in to the national Open Data portal[67], which is since 2017 managed by the Population Register Centre.

However, the Finnish case is not yet the standard in Europe. Although re-users see platforms such as the EDP as an encouraging development, there is still a clear need for centralisation and harmonisation to improve the accessibility of Open Data. It was mentioned that especially on national levels, Open Data is too often scattered around multiple platforms, lacking a central directory[68].

The engagement of policy-makers and politicians in Open Data is another frequently mentioned barrier. Having sponsorship of Open Data initiatives at the highest level could lead to an increase in publishing data at all government levels, creating peer-pressure among administrations. Generating awareness on what can be done with Open Data[69] and its specific economic and societal value[70] helps to convince politicians and policy-makers to prioritise Open Data. Senior leadership is essential for driving change. As this is also an important means to overcome organisational barriers, section 3.2 elaborates further on the role of senior leaders. A good example of a country having sponsorship at the highest level is Slovenia. Data is here not only considered to be a valuable asset for the economy, but also for transforming the public sector and reorganising processes.

Awareness is also crucial for overcoming political barriers on the side of businesses (potentially) using Open Data, as these political barriers relate to the decision-makers that should consider Open Data as a resource. They need to be convinced of the added value of Open Data for them, how it can be integrated in their business model, how it helps to develop new products or services or how it could optimise existing operational processes. Specific measures that can be taken to increase awareness among both data publishers and data users are described in Chapter 3.6.

---

[65] 6Aika website

[66] Helsinki Region Infoshare Portal

[67] National Open Data Portal Finland

[68] European Data Portal, 2017, Re-using Open Data

[69] European Data Portal, 2017, Re-using Open Data

[70] European Data portal, 2015, Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources

3.2 Best practices to overcome organisational barriers

Interaction between the various stakeholders in the data value chain was raised as an important organisational barrier. This refers for instance to interaction between data publishers and data users, but also to interaction between data producers and data publishers. Means to improve the interaction between data publishers and data users are described in section 3.6. As regards the interaction between data producers and data publishers, different publication processes are in place to support the different type of publishers. Issues stemming from this process can be addressed by automating this process as far as possible, as illustrated in the examples below[71].

> **Different type of publishers**
> 1. For **small-scale publishers** with minimal needs, portals owners aim to reduce friction. One provides a username and password for a simple Excel backend. Another allows producers to fill in a simple metadata form directly into their CKAN backend.
> 2. For **medium-scale publishers**, for instance those publishing around 1,000 datasets, portal owners offer more sophisticated processes. One has built an intermediate database where these publishers could manage their own information and metadata, running a script once a week and uploading anything that is new onto the portal.
> 3. For **large-scale publishers** with a professional data need and where data required more regular updating, portals typically have a more automated process. Some national portals fetch data on a daily basis using harvesting nodes.

Institutionalising Open Data, making it part of the daily operations and the daily routines, is another challenge for data publishers. Automated processes, such as described above, help to overcome such issues. National guidelines can be defined to govern the publishing of Open Data. Examples of countries having these national guidelines in place are Italy, Ireland and Slovakia. As these guidelines mainly relate to technical challenges, these best practices are further explored in section 3.4. Teams managing the national Open Data portal often do not have the necessary authority to enforce the publication of public sector information, let alone enforce the quality of such datasets. Therefore, senior leadership is needed (e.g. a Chief Data Officer) in order to drive Open Data within public sector bodies and to enforce standards for the publication of Open Data. The use of senior leaders, such as a Chief Data Officer, proves to be a best practice at the Open Data portals of the United Kingdom and Vienna[72]. The Vienna Data portal has appointed a Chief Data Officer, and he plays a vital role in the roll-out of Open Data in the various government departments of the city of Vienna. Working with the operational part of the ICT-organisation, the Chief Data Officer supports the portal by planning Open Data phase and speaking with departments to help them publish data[73].

Another means to facilitate the institutionalisation of Open Data is by improving the skills of the people that need to work with Open Data.

---

[71] European Data Portal, 2017, Recommendations for Open Data Portals
[72] European Data Portal, 2017, Recommendations for Open Data Portals
[73] European Data Portal, 2017, Recommendations for Open Data Portals

> **Training**
>
> For releasing Open Data, teams managing the national Open Data portal are often dependent on other public bodies and regional governments. These might not all have the technical knowledge to implement for instance features that facilitate automatic uploading and updates of the data. In Greece, the Ministry responsible for the Open Data policy has trained public administration to publish their data and to upload it to the portal. In Spain, in order to overcome the barrier of the lack of technical knowledge among civil servants, a personalised and direct Help Desk for the PSI re-use managers was set up for national, regional and local administrations.
>
> The EDP offers the Training Companion, which helps Open Data teams to deliver training on the basics of Open Data, including training materials[74].

3.3 Best practices to overcome legal barriers

As indicated in section 2.4, an unclear legal framework hampers the opening of data. Data publishers expect and need a framework for publishing their data. Although the PSI directive and its revision set out the general legislative framework, publishers need a clear and practical framework that goes beyond a general transposition of the revised PSI directive. They need specific definitions, roles and responsibilities, as well as accountability for publishing data[75]. Malta and Switzerland are countries that are focusing on providing this framework, as they consider it a key success factor for sustainable Open Data publication.

A second important legal barrier relates to privacy issues. Open Data is deemed privacy sensitive, and hence it prevents the publication of certain data. To overcome this barrier, the Netherlands has adopted an approach in which sensitive datasets are supplied by the national Statistics Office, which only publishes statistical information without identifiable or traceable personal data. For data publishers that wish to promote the use of Open Data while ensuring data controllers' obligation to respect the right of data subjects to personal data protection, a report on Open Data and Privacy[76] is available in which clear guidelines are included.

> **Guidelines on Open Data & Privacy**
>   - Understand the data. Consider potential use cases, the value of the data and potential risks.
>   - Consult. Engage stakeholders (Figure 16) about the publication programme, be aware of additional risks.
>   - Remember the three pillars of privacy, data protection and public confidence.
>   - Be very sure of the grounds for publishing personal data.
>   - Anonymise well and thoroughly. Follow guidelines for anonymising personal data.
>   - Remember utility. There is no point publishing data which has been denuded of serious content.
>   - Don't release and forget. Anonymisation and Open Data are not cheap options.
>   - Have a plan in place in the event of a problem. Be not only transparent, but also

---

[74] European Data Portal, 2016, Training Companion

[75] European Data Portal, Open Data Maturity in Europe 2016

[76] European Data Portal, 2016, Open Data and Privacy

Figure 16: Stakeholders: a five-party dialogue
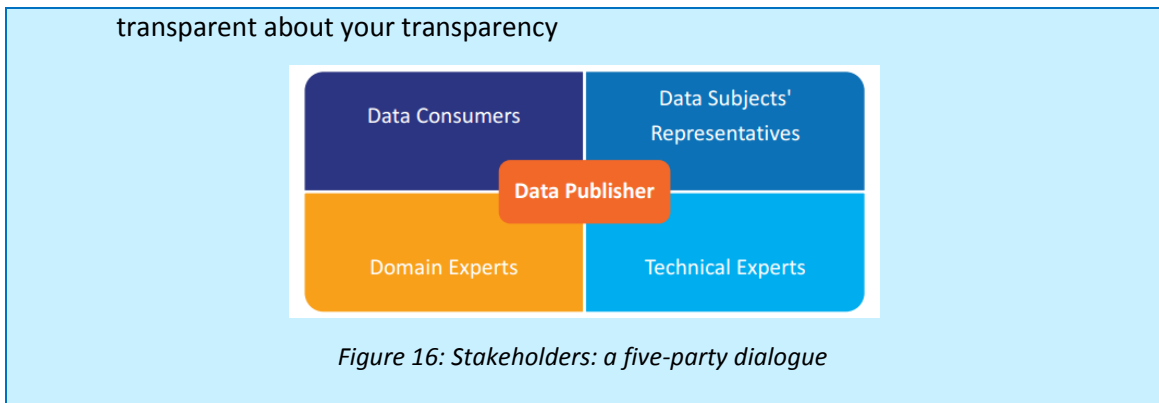
Licences are considered to be another legal barrier. Many different licences are available at both a global and a national level. The EDP currently harvests data carrying 49 different types of licences which range from Open Government and CC-BY licences to specific national licences[77]. To help clarify what can be done with and how the licences can be combined with each other, various tools are available such as the Filter Licence of the EDP.[78]

> **Checking licences**
>
> There are many licences available, from global standards such as the Creative Common (CC) licence (CC 4.0, CC-BY, etc.) to national licences. But how to understand the meaning of all these different licences, and how do you know whether different licences from different datasets are compatible? The EDP offers a Licence Assistant[79] that provides descriptions of the available licences. It also gives an overview of how to apply licences as a distributor of Open Data. For combining Open Data and for checking whether different licences are compatible, the European Data Portal offers the Licence Compatibility Overview[80] also accessible online.

3.4   Best practices to overcome technical barriers

Two important factors currently hampering the re-use of Open Data are the machine-readability of the data and the different formats used to present the data. As shown in section 2.5.1, the amount of machine-readable datasets available on both national Open Data portals and also other portals harvested by the EDP leaves room for improvement. Since Open Data is a new development, Open Data portals have focussed primarily on increasing the availability of data, instead of the quality of the available Open Data. An example of a portal that has all their data available in machine-readable format is the portal of Trentino, Italy[81]. 99% of their datasets are available in an accessible distribution format, meaning that almost all datasets have an AccessURL, according to the DCAT-AP specification: a landing page, feed, SPARQL endpoint or other type of resource that gives access to the distribution of the dataset[82].

---

[77] European Data Portal, Open Data Maturity in Europe 2016

[78] European Data Portal, 2016, Filter licences

[79] European Data Portal, 2016, Licence Assistant

[80] European Data Portal, 2016, Licence Compatibility

[81] Trentino Open Data Portal

[82] W3C, 2014, Data Catalogue Vocabulary (DCAT)

Managing data quality

Ensuring the quality of data is still a problem for many portal owners, and most portals rely on users finding and highlighting errors. Owners of larger portals state that they do not have the capacity for quality control. But there are exceptions. At the United Kingdom Office for National Statistics Portal for instance, quality control managers screen the data uploaded to the portal by publishers. Quality checks for metadata are more common. The Spanish Open Data portal has an automated process in place for detecting metadata errors. This could best be described as a mechanism that informs publishers when metadata does not match with prescribed schemas.

The EDP has its own Metadata Quality Assurance (MQA[83]), monitoring the quality of metadata harvested from other portals. Metadata quality is based on validation against the metadata standard DCAT, and the available distributions of a dataset. The MQA is updated on a weekly basis. Maintenance of existing datasets is an important task for data publishers, to avoid outdated data and broken links. The data maintenance process is described in the European Data Portal Practical Guide[84] (Figure 17).

Figure 17: Data maintenance process

To further develop automated processes, each national portal should have an Application Programming Interface (API) combined with a complete metadata profile. An API allows other tools, such as machines, access to the data on the portal. This allows data users to automatically access the data available on the portal. It saves efforts in manually uploading data and it limits errors when editing data and meta-data manually. The example of Estonia's policy entitled 'API first', underlines how important machine-to-machine communication is for efficient interactions and overall service delivery. This policy dictates that new information systems have to provide access to their Open Data through APIs. Currently, 83% of the EU28+ national Open Data portals have an API[85].

National guidelines can be defined to govern the publishing of Open Data and to ensure data quality. Examples of countries having these national guidelines in place are Italy and Ireland.

---

[83] European Data Portal, 2017, Metadata Quality Dashboard
[84] European Data Portal, 2016, Goldbook
[85] European Data Portal, Open Data Maturity in Europe 2016

> **The use of national guidelines**
>
> An example of solid national coordination is Italy where a metadata application profile has been developed based on the DCAT Application profile. The DCAT-AP_IT serves as a guideline for all public administrations across the country, regardless of the level of government, to comply with when publishing Open Data. To support its implementation, the guidelines were subject to a public consultation during the autumn of 2016. The Open Data portals of the Italian province of Südtirol and Trentino have for instance recently implemented the DCAT-AP_IT profile[86]. Holding public consultations is also a means to drive ownership by giving different stakeholders an opportunity to share their expectations and influence the process at hand.
>
> Ireland has published its Open Data Technical Framework in 2015[87]. This framework supports the ongoing implementation of the Open Data Initiative and ensures that publication of datasets on the Open Data Portal, data.gov.ie, is done in a consistent, persistent and truly open way. It provides a clear set of requirements for public bodies to ensure that published datasets meet clearly defined standards and is interoperable.

A recent report[88] published by the EDP formulates several recommendations that will improve the quality of the (meta-) data. It states that the characteristics of a sustainable Open Data portal's operations should be similar to those of other digital services.

> **Recommendations for sustainable technical operations suggest Open Data Portals should:**
>
> - Have in place automated monitoring systems to ensure that errors, capacity problems and downtime are quickly detected and dealt with
> - Provide clear mechanisms for users to report problems with the service
> - Provide a status page that informs users about known problems with the service
> - Have defined processes and responsible parties to deal with incidents where the service becomes inaccessible for any user
> - Keep track of bugs and feature requests that require development time, preferably through a publicly accessible issue list
> - Have a defined, rapid and preferably automated process for signing new users, particularly publishers, up to the service

Currently, data providers are not prioritising the release of data based on the data that is most re-used. A first step for data publishers to improve the quality of Open Data could be to focus on priority domains. These are the domains most often used. Targeted prime sectors could be defined that should match with user demand. The Bath:Hacked portal is a good example[89]. They ran a series of meetups to discuss different ideas for what to work on and what data could be released. This enabled them to gauge demand within the community, and then use this to leverage data releases. Portal owners can also rely on existing research for identifying priority domains. The European Commission has identified the top five priority data domains, namely statistics, geospatial, transport &

---

[86] Province of Trento, 2017

[87] Irish Department of Public Expenditure and Reform, 2015, Open Data Technical Framework

[88] European Data Portal, 2017, Recommendations for Open Data Portals

[89] Bath:Hacked, 2017

EUROPEAN
DATA PORTAL

infrastructure, companies, and earth observation. These five priority domains overlap to a large extent with the five most consulted domains as indicated by the EU28+[90] (Figure 18).

Figure 18: Top five priority data domains

Focusing more on the needs of re-users could help to address the barrier of the availability of data. This being said, the necessity to provide more information applies in both directions of the data flow. A clear need from re-users of Open Data is to make information on them available. More particularly, it was mentioned that information on the location of businesses and the number of start-ups in a given sector or period is useful[91].

3.5 Best practices to overcome financial barriers

Although the potential financial benefits of Open Data are substantial, as indicated in section 2.6, there are a number of financial barriers that prevent the Open Data community from realising this potential. A loss of revenue was mentioned as being a financial barrier for freely opening up data. Recent research by the EDP indicates that in 2016, 96% of the EU Member States have implemented the marginal cost funding model in law, meaning they can only charge for costs incurred for reproduction, provision and dissemination of Open Data.

Sustainable funding is crucial for Open Data portals to continuously ensure the quality and release of Open Data[92]. First, it is recommended for data publishers to **be open about the funding strategy**. Openness about funding strategies can give the community making use of the portal an opportunity to clarify aspects of the strategy, and share experiences that might help argue for greater funding. Second, when developing a funding strategy or a budget, it is important to **take the lifecycle of a portal into account**, including not only the costs related to the development and launch of the portal, but also one-off expenses and maintaining costs. This is illustrated by Finland, which indicated to

[90] Open Data Maturity in Europe 2016
[91] European Data Portal, 2017, Re-using Open Data
[92] European Data Portal, 2017, Recommendations for Open Data Portals

include factors such as the renewal of the old technical system and the interface in the annual budget. The sustainability report states:

"When a portal is in active use and creating impact, a funding model is likely to be over a number of years with a mix of recurring funding for day-to-day governance, operations and maintenance and one-off funding for developments such as new products and service features or to support updates and enhancements"[93].

Third, a **multiple funding sources model** is considered to be the most sustainable funding model, as this ensures continuity of funding even if one source becomes unavailable. Examples include Austria and Germany. Both national portals have federated financing agreements with multiple regions, based on the size of the population of the regions. Fourth, it is essential that the management team of a portal is able to **draft a clear strategy and budget for the portal**. Open Data portals are a service provided to data publishers and users. As with all services, their owners should have clear budget responsibilities and funding streams. Portal owners need to avoid thinking that Open Data would always be funded as a public good. They need to proactively justify the presence of the Open Data portal, backed by well-documented information about the benefits. This needs to be translated in a clear and up-to-date strategy and budget. This improves the access to sustainable financing.

Lastly, **research on the impact of Open Data** supports the business case for future funding. Impact can be political, social and economic. Political impact refers to government efficiency, effectiveness and transparency, while social impact refers to for instance environmental sustainability and social inclusiveness. Economic impact refers to economic growth, business innovation and job creation[94]. Evidence on the benefits of Open Data in one of these areas supports the claim for funding. Recent research shows that EU member states are increasingly undertaking research activities aimed at measuring the impact of Open Data. An example can be found in Germany, where the Konrad Adenauer Stiftung published a study on Open Data and its benefits. Its shows that Open Government Data in Germany generates an economic value added of 43.1 billion EUR per year and 20,000 jobs[95]. Clearly documented research enables policy makers to justify the presence of the Open Data portal and understand the benefits of publishing Open Data. This is done for instance by Spain.

> **Spain proactively builds the business case for Open Data**
> Spain's national Open Data portal, Datos.gob.es, has played a notably active role in assessing its impact. In 2014, they found that the Open Data infomediary sector was made up of over 350 companies, collectively employing over 4,000 people and generating at least 4,500m EUR[96]. This helped Spain to achieve a top ranking in the 2016 Open Data Maturity index. The Spanish national portal focused on developing value around Open Data, by tailoring their data provision strategy to the specific needs of the user community. Developing evidence of how it has achieved this has been vital for developing the business case for Open Data.

---

[93] European Data Portal, 2017, Recommendations for Open Data Portals, p. 28

[94] Open Data Maturity in Europe 2016

[95] Marcus M. Dapp et al (2016) Open Data: The Benefits. The economic impact for Germany

[96] Data.gob.es, 2014, Characterization Study of the Infomediary Sector

3.6 Best practices to overcome awareness barriers

Evidence[97] shows that awareness of Open Data is an issue among both data publishers and re-users of Open Data. There are still few awareness raising activities around the availability of Open Data, although a growing trend is noticeable. It is important to ensure a continuous dialogue between data publishers and data users, and bringing them together helps to address challenges faced on both sides. One means to do so is to create and leverage on active user communities. User communities, consisting of developers, civil society groups, journalists, political groups, citizens, businesses and academics, not only potentially use the data available on Open Data portals, but they also provide feedback, contribute to online forums and promote the use of data within their networks.

User communities

Some portals are supported by user communities. These communities drive the re-use of Open Data and improvements to both the portal and the quality of the data made available. These user groups can be initiated by (national) governments such as in the United Kingdom, where the national government established the now disbanded Open Data User Group. But user communities also include independent Open Data advocates like the Open Knowledge Network and the Open Data Institute, both having groups and presences in several European countries.

Bringing together publishers and user communities helps to tackle concrete challenges, raises awareness on the side of the data publisher regarding the specific needs of users, and helps portal owners to demonstrate its relevance. Such user events are for instance held in the Netherlands every three months at the Ministry of the Interior[98]. These meetings provide valuable feedback for portal owners on how the portal infrastructure could be improved and what kind of data is most needed. Connecting and maintaining communities around Open Data is essential for continuous feedback. A well-known and often-used means to drive a user community is organising hackathons (see box below). Besides user communities, publisher networks can be set up within countries as part of running a portal. These publisher networks can connect publishers from several governmental departments, and from several government levels (from national to local). It facilitates sharing of experiences and feedback and it promotes best practices, leading to more data and a higher quality of the data published.

Hackathons and other events to drive awareness

Some countries are particularly active in organising events to drive awareness. A hackathon is a sprint-like event in which developers and others involved in graphic design, software development and data management gather together and collaborate intensively on developing software projects. The first edition of #NRWhackathon in Germany for instance focused on developing educational applications, whereas the 'Game of Code 2016' hackathon in Luxembourg focussed on general applications. Estonia is another example of a country being active in organising hackathons, for instance the Garage48 Hackathons[99]. In addition to this hackathon, Estonia also organised the Nordic Digital Day 2016 about the opportunities and

---

97 [European Data Portal, 2017, Re-using Open Data](#)
98 https://data.overheid.nl/bijeenkomsten
99 http://garage48.org/
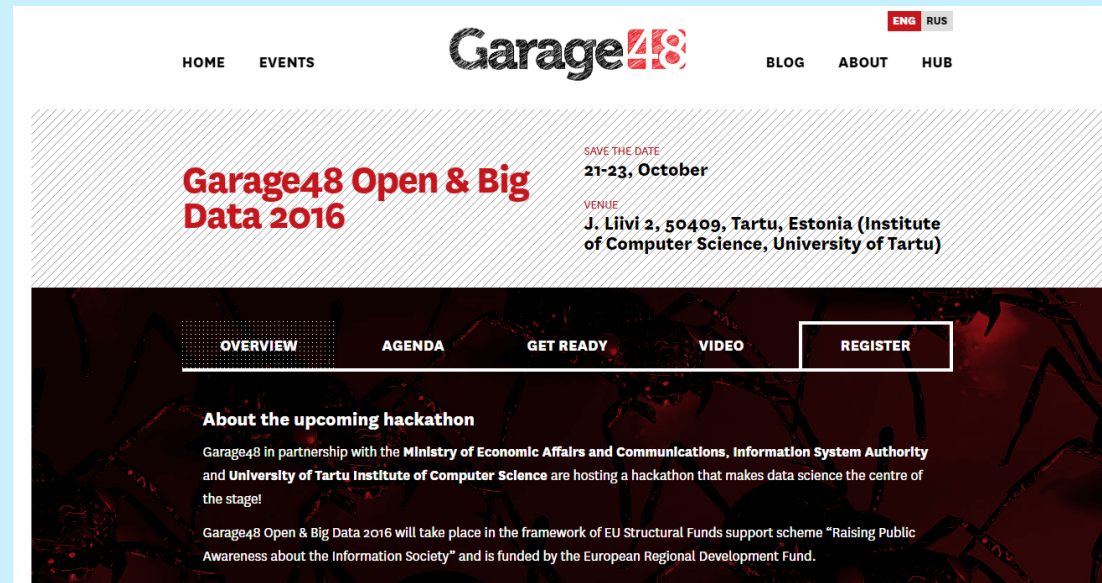
challenges an emerging data society brings.

Figure 19: Garage48 Hackathon in Estonia

Storytelling is an important method to drive awareness on the benefits of Open Data and the specific needs of users of Open Data. Stories make the use of Open Data more tangible and the value more insightful. In addition to the use cases available on the EDP, there are already a number of initiatives across Europe to collect and share stories of Open Data organisations. In Germany for example the website www.datenwirken.de provides an overview of applications based on Open Data. Also many national Open Data portals provide attention to Open Data applications, for instance the Romanian Open Data portal[100] (Figure 20).
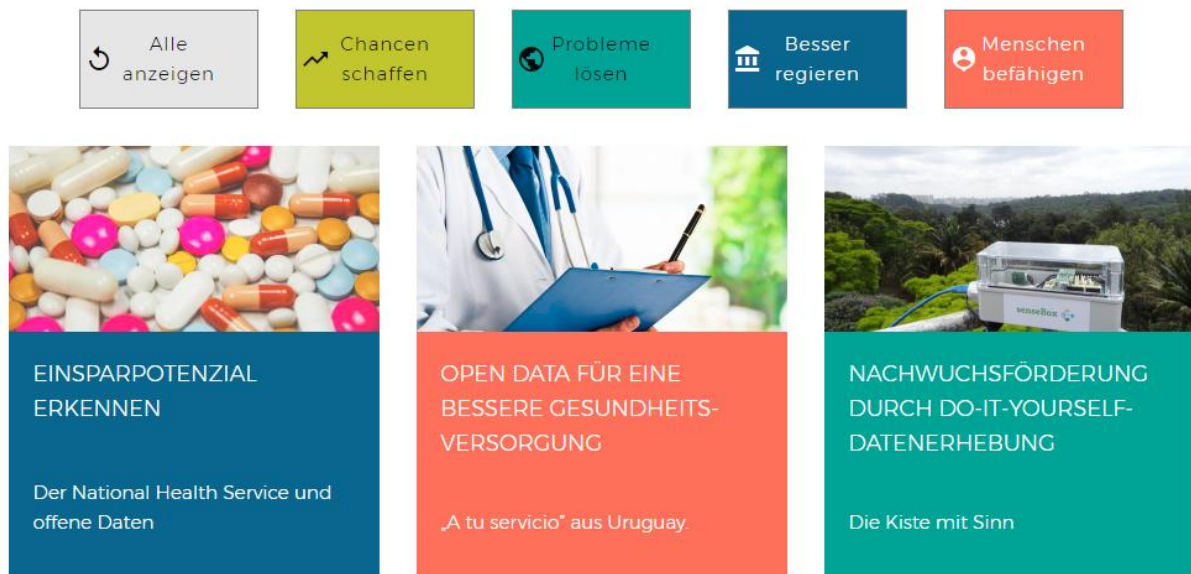

Figure 20: Examples of use cases on www.datenwirken.de

---

[100] http://data.gov.ro/showcase

4 Conclusion

Despite the substantial potential value of Open Data as indicated in several studies, barriers remain while moving forward with Open Data. These barriers are found within the political, organisational, legal, technical and financial domains, while also the awareness on both the availability of Open Data and the specific needs of users of Open Data leave room for improvement. Working with geospatial data in the Open Data field brings some additional barriers on the table, since geospatial data is largely driven by the INSPIRE Directive instead of the (revised) PSI Directive. This results in some gaps, with different standards being used in the geospatial community as compared to the Open Data community.

The barriers are faced by different actors, being either data publishers or data users. The barriers are related and often dependent on each other. The actors involved are often dependent on each other as well, as challenges faced by one actor might have consequences for the other. Although both data publishers and re-users of Open Data have found ways to overcome some of the barriers, solutions are sometimes rather stand-alone, targeting only one specific barrier one has to overcome. More-over, solutions and best practices are not always shared at a national or European level. In light of the development of Open Data policies it is important to realise that there is no single solution that will solve all the challenges. Instead, policies need to address a combination of interwoven solutions to the different types of interdependent challenges.