

# Rethinking the impact of open data

*A first step towards a European  
impact assessment for open data*

**data.europa.eu**

The official portal for European data



Title: Rethinking the impact of open data

Subtitle: Towards a European impact assessment for open data

This study has been prepared as part of data.europa.eu. Data.europa.eu is an initiative of the European Commission. The Publications Office of the European Union is responsible for contract management of data.europa.eu.

For more information about this paper, please contact the following.

**European Commission**

Directorate-General for Communications Networks, Content and Technology  
Unit G.1 Data Policy and Innovation  
Email: [CNECT-G1@ec.europa.eu](mailto:CNECT-G1@ec.europa.eu)

**data.europa.eu**

Email: [info@data.europa.eu](mailto:info@data.europa.eu)

**Authors:**

Charlotte van Ooijen  
David Osimo  
David Regeczi  
Elena Simperl  
Eline Lincklaen Arriëns  
Hidde Holl  
Jochem Dogger  
Laura van Knippenberg  
Myrte ter Horst  
Oscar Corcho  
Sofie Finn Storan

Last update: January 2023

<https://data.europa.eu/>

**DISCLAIMER**

The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use that may be made of the information contained therein.

Luxembourg: Publications Office of the European Union, 2023

© European Union, 2023



The reuse policy of European Commission documents is implemented by Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Unless otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence (<https://creativecommons.org/licences/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

ISBN 978-92-78-43399-4

ISSN 2022.6260

doi: 10.2830/911822

OA-05-22-442-EN-N

# Contents

- Executive summary* ..... 8**
- Introduction*..... 10**
- 1. Literature study and conceptual analysis of the impact of open data* ..... 12**
  - 1.1. Why carry out an impact assessment?** ..... 12
  - 1.2. Review Academic literature** ..... 13
    - Method for selecting articles ..... 13
    - Analysis of academic literature ..... 13
  - 1.3. Impact definitions from national studies** ..... 18
    - Method to find impact definitions from national studies ..... 18
    - Analysis of national studies ..... 18
  - 1.4. Open data impact domains** ..... 25
    - Social impact of open data ..... 25
    - Political impact of open data ..... 25
    - Economic impact of open data ..... 26
    - Environmental impact of open data ..... 26
  - 1.5. Conclusion: considerations for the development of an open data impact assessment methodology** ..... 28
- 2. Impact measurement on national and municipal open data portals*..... 29**
  - 2.1. National open data portals** ..... 30
    - Cyprus ..... 30
    - Germany ..... 32
    - Estonia ..... 33
    - Spain ..... 35
    - Finland ..... 36
    - France ..... 37
    - Croatia ..... 39
    - Ireland ..... 40
    - Lithuania ..... 41
    - Netherlands ..... 42
    - Poland ..... 44
    - Ukraine ..... 46
    - Key findings across national open data portals ..... 48
  - 2.2 Local open data portals**..... 49
    - Barcelona ..... 50
    - Berlin ..... 51
    - Bordeaux ..... 53
    - Dublin ..... 53
    - Eindhoven ..... 54
    - Florence ..... 55
    - Helsinki region ..... 56
    - Lisbon ..... 57

Madrid.....	58
Paris.....	59
Vienna.....	60
Zaragoza.....	61
Key findings across local open data portals.....	62
<b>2.3 Conclusion: considerations for the development of an open data impact assessment methodology.....</b>	<b>63</b>
<b>3. Measuring the impact of open data intermediaries.....</b>	<b>66</b>
<b>3.1 What are open data intermediaries and why do they matter?.....</b>	<b>66</b>
<b>3.2 What research says about downstream open data reuse.....</b>	<b>68</b>
Intermediary positioning in healthy data ecosystems.....	69
Intermediary-facilitated access to data.....	71
The social, political, economic and environmental impact of intermediaries.....	73
<b>3.3 Conclusion: considerations for the development of an open-data impact-assessment methodology.....</b>	<b>77</b>
<b>4 Impact indicators that can be obtained via automated feedback mechanisms.....</b>	<b>80</b>
<b>4.1 Context: automation and open data impact.....</b>	<b>80</b>
<b>4.2 Starting point from previous research on European data portals and an initial set of proposals for automating the calculation of use and reuse indicators.....</b>	<b>81</b>
<b>4.3 A systematic literature review on automation of impact indicators.....</b>	<b>88</b>
Selection of sources.....	89
Analysis of selected papers and proposal of indicators that can be automated.....	90
<b>4.4 Conclusion: considerations for the automation of impact indicators.....</b>	<b>93</b>
<b>5 Conclusions.....</b>	<b>95</b>
<b>Bibliography.....</b>	<b>97</b>
<b>Appendices.....</b>	<b>103</b>
Annex I – Methodology for inventory of national and local open data portals.....	103
Annex II – Results of national portals’ quick scans.....	106
Annex III – Results quick scan local portals.....	109
Annex IV– Impact measurement on national and local open data portals.....	111
Annex V – open data intermediary case study overview.....	131

## List of Figures

FIGURE 1. REUSE CASE SECTION ON THE CYPRIOT NATIONAL OPEN DATA PORTAL .....	31
FIGURE 2 REUSE CASE SECTION ON THE GERMAN NATIONAL OPEN DATA PORTAL.....	33
FIGURE 3 REUSE CASE SECTION ON THE ESTONIAN NATIONAL OPEN DATA PORTAL .....	34
FIGURE 4 REUSE CASE SECTION ON THE SPANISH NATIONAL OPEN DATA PORTAL .....	36
FIGURE 5 REUSE CASE SECTION ON THE FINNISH NATIONAL OPEN DATA PORTAL.....	37
FIGURE 6 REUSE CASE SECTION ON THE FRENCH NATIONAL OPEN DATA PORTAL .....	38
FIGURE 7 REUSE CASE SECTION ON THE CROATIAN NATIONAL OPEN DATA PORTAL.....	39
FIGURE 8 REUSE CASE SECTION ON THE IRISH NATIONAL OPEN DATA PORTAL.....	41
FIGURE 9 REUSE CASE SECTION ON THE LITHUANIAN NATIONAL OPEN DATA PORTAL .....	42
FIGURE 10 REUSE CASE SECTION ON THE DUTCH NATIONAL OPEN DATA PORTAL .....	43
FIGURE 11 REUSE CASE SECTION ON THE POLISH NATIONAL OPEN DATA PORTAL .....	46
FIGURE 12 REUSE APPLICATIONS ON THE INFRASTRUCTURE IMPACT SUBPAGE OF THE UKRAINIAN OPEN DATA PORTAL.....	47
FIGURE 13 API USAGE ON THE NATIONAL OPEN DATA PORTAL OF UKRAINE .....	48
FIGURE 14 REUSE CASE SECTION ON THE OPEN DATA PORTAL OF THE CITY OF BARCELONA .....	51
FIGURE 13 REUSE CASE SECTION ON THE OPEN DATA PORTAL OF THE CITY OF BERLIN.....	52
FIGURE 16 REUSE CASE SECTION ON THE OPEN DATA PORTAL OF THE CITY OF BORDEAUX.....	53
FIGURE 17 TWO METRICS PROVIDED BY THE DATA PORTAL OF THE CITY OF DUBLIN.....	54
FIGURE 18 REUSE CASE SECTION ON THE OPEN DATA PORTAL OF THE CITY OF EINDHOVEN .....	55
FIGURE 19 RANKING OF MOST-DOWNLOADED DATASETS ON THE OPEN DATA PORTAL OF THE CITY OF FLORENCE.....	56
FIGURE 20 REUSE CASE SECTION ON THE OPEN DATA PORTAL OF THE HELSINKI REGION.....	57
FIGURE 21 REUSE CASE SECTION ON THE OPEN DATA PORTAL OF THE CITY OF LISBON.....	58
FIGURE 22 REUSE CASE SECTION ON THE OPEN DATA PORTAL OF THE CITY OF MADRID.....	59
FIGURE 23 REUSE CASE SECTION ON THE OPEN DATA PORTAL OF THE CITY OF PARIS.....	60
FIGURE 24 REUSE CASE SECTION ON THE OPEN DATA PORTAL OF THE CITY OF VIENNA.....	61
FIGURE 25 REUSE CASE SECTION ON THE OPEN DATA PORTAL OF THE CITY OF ZARAGOZA.....	62
FIGURE 26: CALL FOR INPUT TO THE OPEN DATA COMMUNITY.....	104

## Executive summary

This report is the first in a series of four that aims to establish a standard methodology for open data impact assessments that can be used across Europe. This exercise is key because a consistent **definition of the impact of open data** does not exist. The lack of a robust, conceptual foundation has made it more difficult for data portals to demonstrate their value through empirical evidence. It also challenges the EU's ability to understand and compare performance across Member States.

Most academic articles that look to explore the impact of data refer to existing open data frameworks, with the open data maturity (ODM) and open data barometer (ODB) ones most frequently represented. These two frameworks distinguish between different kinds of impact, and both mention **social, political and economic** impacts in particular. The ODM also includes the **environmental** impact in its framework.

Sometimes, these frameworks diverge from the European Commission's own recommendations of how best to measure impact, as explained in specific sections of the better regulation guidelines and the better regulation toolbox. They help to answer a critical question for policymakers: do the benefits provided outweigh the costs of assembling and distributing (open) data? Future reports in this series will further explore how to better align existing frameworks, such as the ODM, with these critically important guidelines.

### The example set by national data portals

While the overall framework for measuring impact continues to improve, national portals have provided fertile ground to explore how open data is being used. The methods that they have used to explore their own performance and impact include:

- describing and analysing good-practice applications;
- soliciting feedback through forms;
- publishing use and user statistics on a dedicated dashboard on metrics such as dataset popularity, thematic distribution of downloads and thematic distribution of reuse cases; and
- measuring the range and update frequency of datasets.

Local and regional portals are similar to national portals in their ways of measuring impact, but often have less information about open data use available online.

### The impact of data intermediaries

Open data intermediaries serve as a bridge between the data portal or the data provider and the data reuser. The primary output of an open data intermediary is thus data and not data products.

Data intermediaries create impact by leveraging various types of capital to both carve out niches in data value chains by collecting or enriching existing niche datasets, and by compensating for deficits in both producer and user capacities. Moreover, they have a valuable role as lobbyists, by convincing the government to make more government data freely available and informing the government which datasets would be most valuable for reusers.

Currently, few robust analyses exist of the role intermediaries play in open data value chains. These analyses are further complicated by the lack of a consistent definition in the literature. At times, the



group of intermediaries also include infomediaries, who are called ‘data reusers’ in this report. Another barrier for the impact analyses is that the actual market size of open data intermediaries is still unknown. The impact of intermediaries can be assessed in a similar way to that of open data portals, with use case repositories and user statistics. However, it remains unclear whether the impact is created because of the open data intermediary or whether the reusers would have found the dataset regardless.

To enable automated measurements, data publishers (such as open data portals) should continue to provide basic metadata related to the number of views and downloads associated to each dataset. Furthermore, they could also provide application programming interfaces (APIs) access to their data, together with metrics related to data usage. The demand side of open data (by intermediaries and reusers) could start to implement tools such as web crawlers to check data usage in academic literature. Additionally, data portals and open data intermediaries should publish the number of downloads and visits for each dataset, the aggregated number of downloads and visits for each dataset on the different portals where it is published, the availability of datasets in other general-purpose or community-specific services, and a range of other variables that could be used for automated assessments.

### **The main challenges for open-data impact assessment**

Two main challenges remain for the creation of an open data impact assessment. First, there is a disconnect between the ways in which open data impact is often defined and how the European Commission strives to do impact analyses. In the better regulation guidelines for impact assessments the European Commission defines social, economic and environmental impact, excluding the political impact dimension that is referred to in open data literature.

Second, impact indicators of open data portals and proposed indicators for open data intermediaries largely rely on **proxies for impact measurement**. The presence of a use case repository or the availability of user statistics is insufficient to measure the overall impact of open data. The main challenge ahead is to find ways to connect these data points to actual impact domains in order to obtain genuine insights about the impact of open data.

## Introduction

There is growing recognition of the fact that (open) data has significant social, economic and environmental value for different institutions and sectors. To exploit this value and create impact, more and more countries and municipalities are publishing parts of their datasets on their national and local portals, allowing citizens to reuse them for various purposes. While the number of open government data initiatives has increased considerably over the past decade, the impact of these initiatives remains uncertain.

There have been few attempts to establish an objective way to measure of open data impact, and this lack of conceptualisation and method has resulted in a scarcity of empirical evidence on the impact of open data. With anecdotal evidence, any attempts to analyse impact are not fully adequate. Notwithstanding, there exists a broad scale of reports, cases and scientific articles that use the term 'open data impact' in a wide variety of ways. Case studies and theoretical analyses based on qualitative observations are commonly used in scientific literature to examine impact. Analysing such information can help scholars and other practitioners to identify benefits and perform open data policy reflections. These programmes have considerable promise but have often been met with obstacles in turning data collecting into data utilisation with meaningful benefits (Chu and Lee, 2019). Data is frequently hidden and dispersed, with information on the same issue spread through several departments or levels of government, and presented in a variety of forms.

This report will be the first step in the process of defining the methodology for a pan-European study on the impact of open data, and is the first in a series of four.

1. **Survey of publicly available studies on the economic impact of reuse of public data resources.**
2. Draft impact assessment methodology.
3. Applying the new methodology to assess the impact of the reuse of public data resources.
4. Interactive monitoring tool.

To that end, this study provides a literature review of the impact of public data resources (otherwise known as open data) and aims to increase awareness and build knowledge among policymakers, data providers and reusers about the impact of public data resources in Europe. The study is divided into four sections.

- Section 1 is a **conceptual analysis** of the social, political, economic and environmental impact of open data.
- Section 2 provides an **inventory** of the European indicator landscape and approaches to measurement. A scan of national data portals and those of municipal data portals will complement the literature research on existing indicators.
- Section 3 adds to the inventory of indicators found on open data portals and highlights outcome indicators measuring the importance of **downstream reuse through intermediaries**, such as those presented in the infomediary sector reports produced by ASEDIE <sup>(1)</sup>, which track the progress over time of companies with a data-based business model, using quantitative business metrics to evaluate economic value.

---

(1) <https://www.asedie.es/en/annual-report>

- Section 4 lists outcome indicators that have been generated through **automated feedback mechanisms**. This includes the tracking of dataset usage and reuse through systematic data citation and tracking mechanisms (e.g. those provided by European Open Science Cloud core services), along with other measurements based on application programming interface (API) call logs for dynamic data.

# 1. Literature study and conceptual analysis of the impact of open data

As a first step in the conceptual analysis of open data impact, this section describes the current available insights regarding the different definitions of impact or value that is created with open data.

Section 1.1 explains why an open data impact assessment is a worthwhile exercise and what guidelines the European Commission offers for doing impact assessments. Section 1.2 provides an overview of academic articles that consider open data impact between 2019 and May 2022. Specifically, this section looks at the different definitions of impact and the multitude of ways in which an open data impact framework can be constructed. Section 1.3 gives an overview of available data impact assessments carried out by national governments and information on the way European countries define and measure the impact of open data. For the latter, input from *the Open Data Maturity Report 2021* <sup>(2)</sup> is used. Section 1.4 provides an overview of the most common impact domains in open data and provides uses cases for these impact domains.

## 1.1. Why carry out an impact assessment?

Measuring impact provides an understanding of both the positive and negative effects of an intervention, in one or several domains. In the light of the open data directive <sup>(3)</sup> and the Implementing Act on High Value Datasets <sup>(4)</sup>, measuring the impact of open data is a worthwhile exercise. While countries are spending a lot of effort to create open data portals to facilitate reuse, the question remains: what is the actual impact of open data, and what domains does it concern?

Nevertheless, before being able to measure impact, we must have an idea of what impact is. Impact can be created in a myriad of ways in an endless number of domains, which makes measuring impact a complex task. It's not always easy to put a number on impact, especially when several domains are impacted at the same time. How should we weigh economic benefits against environmental or health-related downsides?

In 2021, the European Commission published its better regulation guidelines <sup>(5)</sup> and toolbox <sup>(6)</sup>, which address the need for impact assessments to assess the merits of policies. According to the better regulation guidelines, 'impact assessment promotes informed decision-making and contributes to "better regulation" that delivers the full benefits of policies at minimum cost, while upholding the principles of subsidiarity and proportionality'.

The European Commission's stance on how impact assessments should be conducted, as described in the better regulation guidelines, for example, is that impact assessments should focus on all social, economic and environmental impacts. The better regulation guidelines provide a short description for each of the three impact domains.

---

<sup>(2)</sup> Specifically question 35 of the impact dimension: 'Have there been any studies conducted in the past year that focus on assessing the impact of open data (whether political, social, environmental or economic)?' (<https://data.europa.eu/en/dashboard/2021>).

<sup>(3)</sup> <https://digital-strategy.ec.europa.eu/en/policies/legislation-open-data>

<sup>(4)</sup> <https://digital-strategy.ec.europa.eu/en/news/commission-seeks-views-implementing-act-high-value-datasets>

<sup>(5)</sup> Better regulation guidelines: [https://ec.europa.eu/info/sites/default/files/swd2021\\_305\\_en.pdf](https://ec.europa.eu/info/sites/default/files/swd2021_305_en.pdf).

<sup>(6)</sup> Better regulation toolbox: [https://ec.europa.eu/info/sites/default/files/br\\_toolbox-nov\\_2021\\_en\\_0.pdf](https://ec.europa.eu/info/sites/default/files/br_toolbox-nov_2021_en_0.pdf).

- **Social.** Requirements linked to the promotion of a high level of employment, the guarantee of adequate social protection, the fight against social exclusion, and a high level of education, training and protection of human health are relevant in defining social impacts.
- **Economic.** The costs and administrative burden that businesses bear; the impact on small and medium-sized enterprises (SMEs); the effects on sectoral competitiveness, trade and investment flows; the functioning of the internal market and competition; and the decrease or increase of public budgets are taken into account when measuring economic impact.
- **Environmental.** The better regulation guidelines provide a non-exhaustive list of topics that might be of interest in assessing environmental impact: air quality, water quality and quantity biodiversity, soil quality or resources and land use change or degradation, waste production and recycling, zero pollution and toxicity, efficient use of resources (renewable and non-renewable), contribution to circular economy, the likelihood or scale of environmental disasters, and international environmental impacts.

This is not to say that impact must necessarily be measured across these three domains. The remainder of this section will demonstrate that environmental impact is currently not a part of open data impact domains, whereas many domains have added political impact as third dimension.

## 1.2. Review Academic literature

### Method for selecting articles

In order to select academic articles that define the impact of open data, a systematic search methodology is applied. This methodology provides a structured, systematic and transparent manner for selecting articles.

As the first step of the search, the search query ‘Open data impact’ was entered into Google Scholar with a filter that excludes all articles from before 2019. This search query yielded 158 results. A first scan of these 158 articles based on their title and abstract resulted in a subset of 65 articles. The 65 articles were filtered for any duplicates and then scrutinised for relevance, according to the following criteria.

- The language must be English.
- The journal in which the research is published must be included in an official university library.
- The article must be in the range of open data impact or using a definition of open data impact in order to be selected.
- Articles focused on individual impact, such as personal skills, are not selected.

First search query N = 158	First scan based on titles and abstract. N = 65	First subset after scrutinising articles for relevance N = 21	Second search – duplicates – not relevant + first subset N = 25
-------------------------------	--	--	--

Subsequently, two additional queries were performed on ‘defining open data impact’ and ‘impact of open data’. This added four more articles to the sample, bringing the sample to 25.

### Analysis of academic literature

Assessing the literature on the impact of open data shows that open data itself is a contested concept. In many articles open data seems to be equated with open government data, even though these are not necessarily the same concepts. Moreover, open data is often interpreted as machine readable and easily accessible. While this is preferable, it is not necessary for information to constitute open data.

Out of the 25 relevant articles, a differentiation can be made between 1) research that uses an existing framework, 2) literature that tries to develop a definition of its own of the impact of open data, and 3) literature that did not provide any definition for open data impact but did shine a light on how open data has an impact on society through specific use cases. Since this subsection focuses on the development of an open data impact framework, we will consider all articles that either use an existing framework or provide a new definition. We will first explore the three most commonly used frameworks and hereafter focus on the articles that provide their own definition for the impact of open data.

### Articles using an existing framework

Nine of the 25 selected articles use an existing framework. The three frameworks that are referred most often to are: the open data maturity (ODM) report, the open data barometer (ODB) and the open data framework by Verhulst and Young (2017) for open data in developing countries. For each of these frameworks, a short description is provided below.

#### The open data maturity report <sup>(7)</sup>

The ODM is a yearly report published by data.europa.eu that compares the maturity of open data in European countries <sup>(8)</sup>. It evaluates the EU Member States and a number of additional countries: Georgia, Iceland, Montenegro, Norway, Switzerland, Ukraine and the United Kingdom.. The ODM looks at open data impact from four angles.

- **Social impact.** ‘Assesses the extent to which open data has an impact on societal challenges, such as the inclusion of marginalised groups in society, raising awareness on housing in urban areas, and health and well-being related issues’.
- **Political impact.** ‘Focuses on the benefits that open data has in three domains: improving government efficiency, improving government effectiveness, and increasing transparency and accountability’.
- **Economic impact.** ‘Considers aspects such as macro- and micro-economic impact, economic benefits for public administrations’.
- **Environmental impact.** ‘Considers aspects such as raising awareness on the water and air quality, noise levels in cities, waste management systems, environmental-friendly transport systems’.

#### The open data barometer <sup>(9)</sup>

The ODB was published yearly by the World Wide Web foundation and the Omidyar Network until 2018 <sup>(10)</sup>. The goal of the barometer was to assess how governments publish and use open data for

---

<sup>(7)</sup> Used in Benedetti et al. (2020), Davies et al. (2019), Runeson, Olsson and Linåker (2021) and Park and Gil-Garcia (2022).

<sup>(8)</sup> <https://data.europa.eu/en/dashboard/2021>.

<sup>(9)</sup> Myeong et al., 2021.

<sup>(10)</sup> <https://opendatabarometer.org/>

accountability, innovation and social impact. The barometer looks at 30 countries spread over all continents. The ODB makes the distinction between social, political and economic impact.

- **Social impact.** Considers the impact on environmental sustainability and the inclusion of marginalised groups in policymaking.
- **Political impact.** Concerns government efficiency, transparency and accountability.
- **Economic impact.** Assesses whether there is a noticeable positive impact on the economy and whether entrepreneurs successfully use open data to innovate and create new businesses in their countries.

### **The impact of open data on developing economies <sup>(11)</sup>**

Verhulst and Young (2017) wrote a book about open data in developing economies, in which they also developed an impact taxonomy. The authors describe impact domains differently than the ODB and the ODM. More than providing a manner to measure the impact, they describe how open data has an impact. The authors use the four impact categories described below.

- **Improving governance.** Open data can be used for greater transparency, more citizen involvement, better and more efficient service delivery.
- **Empowering citizens.** Open data leads to more information for citizens and therefore improved decision-making capacity and choice.
- **Creating economic opportunity.** The accessibility of open data can enable business creation, foreign investment, meaningful job creation and opens the door for frugal innovation efforts in the public sector.
- **Solving public problems.** Open data can help improve situational awareness in crisis situations and the accessibility of open data can help to bring a wider range of expertise and knowledge to find solutions to public problems.

Although Verhulst and Young describe the impact domains in a different way than the ODB and the ODM, there are some similarities. ‘Creating opportunity’ as described by Verhulst and Young falls nicely in line with what the ODM and ODB call ‘economic impact’. ‘Improving governance’, ‘empowering citizens’ and ‘solving public problems’ can be classified as impact in either the social or the political domain. Note that environmental impact is not mentioned in the book by Verhulst and Young.

The three studies follow a specific pattern regarding the impact domains: all three assess the impact of open data through a social, political and economic lens. The environmental impact is mentioned in both the ODM and the ODB, but the ODM uses it as a separate impact domain, whereas the ODB includes the environmental impact in the social domain. Other recurring topics are the impact that is created with open data through transparency of government information and innovation in the public sector.

### **Articles that aim to develop an own definition of open data impact**

Out of the selected articles, seven develop their own definition of open data impact. Table 1 provides a summary of these articles and their definitions of open data impact. The table also indicates whether the definition encompasses the same categories as seen in the open data frameworks, i.e. social

---

<sup>(11)</sup> Used in Saxena (2020) and Neves, de Castro Neto and Aparicio (2020).

impact, political impact, environmental impact (including gains in government transparency) and economic impact.

Regarding the definition of open data impact, the following stands out. The definition provided by Kawashita, Baptista and Soares (2020) is similar to that of the ODB: 'Open government data impact investigates what economic, social, political benefits open government data initiatives might generate'. Shahrudin (2020)- has a similar definition as well: 'open government data could generate not only economic value but also social value in terms of improved governance and knowledge production' <sup>(12)</sup>. Moreover, all articles that provide their own interpretation of open data impact cover social impact in some shape or form. Krishnamurthy and Awazu (2016) provide a definition of open data impact that mainly highlights the social and political impact open data creates. In their definition, open data 'increases transparency, promotes participation and fosters collaboration'.

Interestingly, none of the academic articles that do not refer to an existing open data impact framework include environmental impact in their definition. Yet, many use cases have an environmental impact (and some will be discussed later in this section). The impact of open data on the environment seems under-conceptualised in contemporary literature, although it could indeed be argued that environmental impact is not the main impact area of open data.

All seven articles refer to the social impact of open data, with some highlighting the positive effects open data can have on quality of life, for example Zuiderwijk and Janssen (2014). Economic impact, through innovation or job creation for instance, is mentioned in six out of seven articles, whereas political impact through transparency or improved governance is mentioned in five out of seven articles.



Table 1: Summary of the articles that do not use an existing framework to define open data impact

#	Article	Clear mentions of impact	Social	Political	Economic	Environment
1	Open data policies, their implementation and impact: a framework for comparison (Zuiderwijk and Janssen, 2014).	'Harrison et al. (2012) identify <b>seven types of impacts</b> , namely financial, political, social, strategic, quality of life, ideological and stewardship.'	x	x	x	
2	Liberating data for public value: the case of Data.gov. (Krishnamurthy and Awazu, 2016).	'The <b>impact</b> of open data initiative <b>on delivering its intended promise</b> : increasing transparency, promoting participation, and fostering collaboration.'	x	x	x	
3	Socio-economic effects and the value of open data: a case from Sweden (Apanasevic, 2021).	'Cost-Benefit-Analysis aims to assess open data impact in <b>monetary value</b> (Koski, 2015). At organisational level, costs of open data could be classified.'	x		x	
4	Guidelines for analysing pathways to impact: evaluation of open data for development (Lokers, Miguel Ayala and Berdou, 2019)	'Open data efforts do not seem to follow a clearly defined path aimed at achieving impact and generally <b>no systematic monitoring of effects</b> is in place'	x	x		
5	An assessment of open government data benchmark instruments (Kawashita, Baptista and Soares, 2020).	'Open government data impact investigates what <b>economic, social, political benefits</b> open government data initiatives might generate.'	x	x	x	
6	Open government data: development, practice, and challenges (Shaharudin, 2020).	'Open government data could generate not only <b>economic value</b> but also <b>social value</b> in terms of improved governance and knowledge production.'	x	x	x	
7	A research agenda on open data impact process for open innovation (Corrales-Garay, Ortiz de Urbina Criado and Mora-Valentín, 2020).	'the fourth phase (Impact) addresses the effects of reusing open data and the <b>innovation</b> that has been created.'	x		x	

### 1.3. Impact definitions from national studies

#### Method to find impact definitions from national studies

The ODM report evaluates whether countries have impact definitions for open data in place. The ODM uses a questionnaire to assess open data maturity within Europe. The questionnaire is filled in by country representatives, who are required to deliver evidence to support their statements. To assess whether countries have a definition, the answers provided to the following four questions are assessed.

- Has your government specified what ‘impact of open data’ means (e.g. in a strategy document)?
- Do you have a methodology in place to estimate the impact of open data in your country?
- Have you or other public bodies launched or performed any activities in the past year to assess the social, political, economic and environmental impact of open data (such as systematic monitoring, commissioning studies, surveys)?
- Is there collaboration between government and civil society or academia to create open data impact (whether social, political, economic or environmental)?

#### Analysis of national studies

Table 2 shows how the countries that participate in the ODM defined the impact of open data. Note that some did not yet have a definition, while others refer to definitions provided by other organisations, such as data.europa.eu and the ODM. Also indicated in the table is whether the open data impact definition used by the countries considers social, political, economic and/or environmental impact and whether public bodies launched or performed any activities in the past year to assess the impact of open data.

Most of the countries have a **holistic view** on the impact of open data, considering the social, political, economic and environmental impact. Some examples are the following.

- **Cyprus.** ‘By “open data impact” we mean all those changes, improvements, opportunities, direct and indirect, which occur through the reuse of open data. This impact is not limited to a single area and can take many forms: economic, social political, environmental, etc.’
- **The Netherlands.** ‘Societal and social impact (i.e. influence of open data on society), economic impact (i.e. the influence of open data on value development for citizens, entrepreneurs or the government itself), environmental impact (i.e. the influence of open data on the environment and climate), and political and administrative impact (i.e. the influence of data use on the efficiency of government services)’.
- **Slovenia.** ‘Open Data have impacts primarily on four main areas: 1) transparency of the work of public institutions (anti-corruption; political impact), 2) innovative digital economy, 3) solving environment related problems (smart cities), 4) efficient functioning and data-driven decisions of the public institutions.’

However, some other countries apply a **narrow** definition. Denmark for instance, defines the impact of open data as ‘the value of the effect on production and efficiency in the private and in

the public sector’, and Finland uses the following definition: ‘the aims are to support public welfare and to provide business opportunities’.

In comparison with the academic literature, national governments seem more appreciative of the positive impact that open data can have on the environment, for instance in reducing air pollution. Innovation, on the other hand, is less often named as a part of open data impact.

Some countries put **key performance indicators** (KPIs) in place to measure the impact of open data. For instance, Austria mentions these five KPIs for the impact of open data on their portal:

1. the number of visitors,
2. the number of downloads,
3. the demand and frequency of use of open data,
4. the number of data providers using open data,
5. the number of applications that have arisen from the use of the data.

The five indicators mentioned by Austria are also mentioned in an earlier report from data.europa.eu on open data best practices <sup>(13)</sup>. Note that these KPIs help to track the use of open data, but that open data use is not equal to open data impact. Hence measuring impact without an established relationship between one of the KPIs and actual social, political, economic or environmental impact is a challenging (if not impossible) task.

---

<sup>(13)</sup> [https://data.europa.eu/sites/default/files/analytical\\_report\\_16\\_top-performing-countries.pdf](https://data.europa.eu/sites/default/files/analytical_report_16_top-performing-countries.pdf)

Table 2: Definitions for the impact of open data provided by country representatives for the ODM study

Country	Impact definition	Report	Social	Political	Economic	Environment
Bulgaria	'Open data impact covers the activities carried out to monitor and measure open data reuse and the impact resulting from the reuse in the <b>political, social, environmental and economic</b> areas. This impact increases transparency, accountability, government efficiency, government effectiveness, inclusion of marginalised groups, raising awareness and can help address a number of issues in various areas.'	<a href="#">Bulgaria – Open data – e-gov report</a>	x	x	x	x
Czechia	'The impact of open data is defined in the methodology for calculating the impacts of open data in Czechia. Since 2017, <b>a study of the impacts of open data in Czechia has been carried out every year</b> . The results of the study were regularly published in the annual report on the state of open data in Czechia, where the methodology for calculating the impacts of open data is also published.'	<a href="#">Czech – Open data</a>				
Denmark	'The impact is defined as the value of the <b>effect on production and efficiency</b> in the private and in the public sector.'	<a href="#">Denmark – Open data – Erhvervsstyrelsen</a>			x	
Germany	'There is <b>no one-size-fits-all definition</b> of the impact of open data in Germany (which would, due to the constitutional division between federal level (Bund) and regional level (Länder) as mentioned above, hardly be applicable to Germany): most public bodies and institutions dealing with open data have delineated what the impact of open data (can) entail from their perspective.'	<a href="#">Germany – Open data – ergebnisse-der-oeffentlichen-konsultation-data</a>	x	x	x	x
Estonia	A 'data-driven state where <b>any decision is based on data</b> and that is reached by re-using data'.	<a href="#">Estonia – Open data – open-data-market-size-in-estoni</a>	x	x	x	x
Ireland	'A <b>macro</b> impact evaluation to examine the broad outcomes of the initiative from a <b>social, political and economic</b> perspective. At a <b>micro</b> level, particular case-studies can be explored to get a clear understanding of the impact of Open Data <b>in specific sectors and under a certain set of conditions</b> .'	<a href="#">Ireland – Open data – Evaluation-Framework.</a>	x	x	x	

Country	Impact definition	Report	Social	Political	Economic	Environment
Greece	'According to the European Data Portal, the OECD's OURIndex, OGP's IRM and the WWW Foundation guidelines. A <b>robust and universal</b> method is still needed.'	<a href="#">Greece – open data-report</a>	x	x	x	x
Spain	' <b>Any positive effect or benefit</b> obtained directly or indirectly <b>for individuals, communities or society as a whole</b> , which occurs over a certain period of time and which results from the development of different activities in a given area characterised by the use of open data as a means to an end.'	<a href="#">Spain – Open data – ferrer –</a>	x	x	x	x
France	'The final report of the Open Data Mission led by MP Éric Bothorel includes a recommendation to evaluate the economic, social, and scientific impact of the opening and sharing of data and source codes. The mission also stresses four impact areas of open data: <b>scientific</b> as a vector of knowledge, <b>economical</b> as a driver of innovation, <b>democratic</b> to improve public service, and <b>political</b> to restore people's confidence in public action.'	<a href="#">France – Open data – Mission_Bothorel_Report</a>	x	x	x	x
Croatia	'The <b>ODM</b> provides the <b>benchmark regulations</b> regarding open data sharing.'	<a href="#">Croatia – Open data – standardi-i-prirucnici</a>	x	x	x	x
Italy	'There are <b>many official definitions</b> of open data. Some examples, starting from the National Guidelines for the valorisation of public information heritage.'	<a href="#">Italy – Open data</a>				
Cyprus	'By 'open data impact' we mean all those <b>changes, improvements, opportunities, direct and indirect</b> , which occur through the reuse of open data. This impact is not limited to a single area and can take many forms: <b>economic, social, political, environmental</b> , etc.'	<a href="#">Cyprus – Open data – Impact report</a>	x	x	x	x

Country	Impact definition	Report	Social	Political	Economic	Environment
Latvia	‘In addition to opening up data for use by <b>industry, researchers and citizens</b> in general, one of the main reusers of data is <b>public authorities</b> themselves. Public authorities can use data to improve the <b>efficiency of the authority</b> , to provide <b>better services</b> , to develop <b>legislation based on data (research)</b> , and to increase <b>citizen involvement in decision-making</b> . The use of open data for in-depth research, monitoring or forecasting offers the potential for <b>good governance</b> and can contribute to <b>economic and social development</b> in regions.’	<a href="#">Latvia – Open data</a>	x	x	x	x
Lithuania	‘Increased <b>transparency in decision-making</b> processes in Lithuania’	<a href="#">Lithuania – Open data – methodology</a>	x	x	x	x
Luxembourg	Built upon open data, reuses are meant to leverage a new digital data ecosystem providing straightforward added value for <b>data producers</b> as well as <b>long term economic return</b> . An important point of the strategy implies the promotion of emerging businesses specialized in data handling and data science.	<a href="#">Luxembourg – Open data – data.public</a>	x	x	x	x
Hungary	‘Making public data available will kick-start the domestic data industry, which can contribute billions of forints to <b>domestic GDP</b> and <b>create thousands of jobs</b> .’	<a href="#">White paper Hungary 2016</a>	x		x	
The Netherlands	‘Impact stands for the extent to which something influences the <b>effect in processes</b> . It is relevant for data providers to know the impact of the data they make available. The impact can be broad. We distinguish the following types of impact: <b>Societal and social</b> impact; the influence of open data on society; <b>Economic</b> impact; the influence of open data on value development for citizens, entrepreneurs or the government itself; <b>Environmental</b> impact; the influence of open data on the environment and climate; <b>Political and administrative</b> impact; the influence of data use on the efficiency of government services.’	<a href="#">The Netherlands – Open data – wat-is-data-impact-en-hoe-meet-je-het</a>	x	x	x	x

Country	Impact definition	Report	Social	Political	Economic	Environment
Austria	‘Studies provide a comprehensive overview of the <b>economic, social and ecological</b> added value of using data. Organisations are encouraged to commission studies on data reuse. Current <b>KPIs</b> are: 1. the number of <b>visitors</b> , 2. the number of <b>downloads</b> , 3. the <b>demand and frequency</b> of use of open data, 4. the <b>number of data providers</b> using open data, 5. the <b>number of applications</b> that have arisen from the use of the data.’	<a href="#">Open Data Analyse – Bessere Entscheidungen treffen</a>	x		x	x
Poland	‘In the Open Data Program, the impact of open data was defined by the <b>benefits</b> its implementation will bring <b>to different groups of stakeholders.</b> ’	<a href="#">Poland – Open Data – raport-z-badania</a>	x		x	
Romania	‘In other words, impact can be defined as the <b>change in the final situation compared to the initial situation</b> of the studied object due to the studied intervention.’	<a href="#">Romania – Open data</a>				
Slovenia	‘Open Data has impact primarily on 4 main areas: 1) <b>transparency of the work of public institutions</b> (anti-corruption; political impact), 2) <b>innovative digital economy</b> , 3) <b>solving environment related problems</b> (smart cities), 4) <b>efficient functioning and data-driven decisions of the public institutions.</b> ’	<a href="#">Slovenia – Open data – Economic_impact_of_open_data_in_Slovenia</a>	x	x	x	x
Finland	‘The aims are to <b>support public welfare</b> and to <b>provide business opportunities.</b> ’	<a href="#">Finland – Open data – report</a>	x	x	x	x
Sweden	‘There is <b>not one strict definition</b> of ‘impact of open data’ that is used in harmonized manner, rather there are <b>several descriptions of various levels of details</b> that outlines the concept of impact.’	<a href="#">Sweden – Open data – oppna-data-datadriven-innovation-och-ai</a>	x	x	x	x

Ukraine	<p>‘For <b>micro studies</b> that we currently conduct to assess the impact of open data in particular sectors we usually follow this <b>KPIs – number of views of the dataset – number of services and products that were created or additional functions were created</b> or integrated the dataset into the existing product – <b>number of users(citizens)</b> of these services. If there is high use of services by citizens it shows that the dataset solves particular problems for them. For <b>macroeconomic</b> impact study that was conducted the following metric were calculated – the <b>rise in GDP, economic gain</b> in UAH and <b>creation of jobs</b> was calculated.’</p>	<a href="#">Ukraine – Open data – dataset/old_result_vnd</a>	x	x	x	x
United Kingdom	<p>‘Open data has <b>different impacts in different contexts</b>. It should be noted that the UK does measure impact on a different level than the EU does. Environmental issues are more specified within geographical data.’</p>	<a href="#">UK – Open data – opengovernment.org</a>	x	x	x	



## 1.4. Open data impact domains

The analysis of the academic literature and national studies clearly shows the need for a common definition of open data impact that can be used consistently. The social, political, economic and environmental categories are commonly used, but there are differences in what these categories entail. This section summarises insights from the academic literature and the answers provided by country representatives for the ODM. Additionally, good practices and use cases of impact creation through open data per impact domain are provided.

### Social impact of open data

Social impact can be evaluated by observing how research and policies affect society as a whole, and how they can help address societal challenges, for example by fostering the inclusion of marginalised groups in society, finding solutions to housing crises in urban areas and raising awareness on the importance of health and well-being (namely the concept of quality-adjusted life years), especially in the context of the COVID-19 pandemic.

#### Use cases that highlight social impact

An example of a service that uses open data to raise awareness on housing in urban areas is the **Spanish** web application Inspide<sup>(14)</sup>. This application uses open data to show whether a city, such as Madrid, adapts to the current requirements of social distancing. To do this, the application shows the width of the pavements represented with colours and provides figures indicating whether their width is sufficient to maintain the requirement distance between pedestrians.

Another example of a country that uses open data to support raising awareness on health and well-being is **Romania**. Following the spread of COVID-19, the Romanian government created an open-data-based national dashboard named 'Monitorizare Coronavirus'<sup>(15)</sup>. Examples of available datasets on the dashboard are the number of registered positive cases, registered deaths linked to COVID-19, the number of patients admitted to the ICU and the number of vaccines performed, including which vaccine provider.

### Political impact of open data

Political impact can be created through better governance, increased transparency of governments, better quality of information available for citizens and less corruption. Trindade et al. (2020) wrote that open data 'helps to solve complex public problems, improves governance and empowers citizens'. The Bulgarian definition notes that open data can increase transparency, accountability, government efficiency and government effectiveness.

#### Use cases that highlight political impact

An example of a service that uses open data for better governance is the **Slovenian** application Erar<sup>(16)</sup>, developed in 2013 by the Commission for Prevention of Corruption<sup>(17)</sup>. This application provides the public with free-of-charge and user-friendly access to information on the business transactions of public sector bodies. Erar's system increases the level of responsibility of public-office holders for

---

<sup>(14)</sup> <https://distanciamiento.inspide.com/>

<sup>(15)</sup> <https://coronavirus.casajurnalistului.ro/en/>

<sup>(16)</sup> <https://erar.si/>

<sup>(17)</sup> <https://www.kpk-rs.si/en/>

effective and efficient use of public finance, decreases the risk of illicit management of funds and limits systemic corruption in Slovenia.

**Lithuania** is another country that uses open data for political impact, more specifically for transparency and accountability. In 2020, the Lithuanian Public Procurement Office (PPO) <sup>(18)</sup> published data on public contracts regarding the supplies and services aimed at combating COVID-19 under an open license. This includes several types of purchases, including protective equipment, testing materials, and transportation services, and data on the buyer, supplier, closing of contracts and prices <sup>(19)</sup>.

## Economic impact of open data

The economic impact of open data can be evaluated at the macro or micro level. An example of macroeconomic benefits would be the growth of the gross domestic product of a region or country. At the micro level, the impact could be assessed by taking a look at the revenue raised by large companies and SMEs. Open data enables organisations to start innovative new business models, which might also lead to macroeconomic benefits for society as a whole in the long run.

Another important and recurring theme in the literature on the economic impact of open data is the benefits that could materialise for developing countries with open data. Within the scientific literature, developing countries are often used as case studies of economic growth through open data and data sharing.

### Use cases that highlight economic impact

One of the countries that uses open data at a macroeconomic level is **Estonia**. Statistics Estonia <sup>(20)</sup> and the Estonian Ministry of Social Affairs <sup>(21)</sup> developed an application that allows one to visualise labour policy indicators between 2005 and 2020. The data used to calculate the indicators are gathered in the Estonian Labour Force Survey <sup>(22)</sup> and are available to download for free from the application's website.

Another example of a country that uses open data at a microeconomic level is **Belgium**. The Belgian start-up Mediafin <sup>(23)</sup> created the online tool called 'Open the box' <sup>(24)</sup>, a service that allows users to discover business networks in specific fields. 'Open the box' allows users to look inside the web of connections between companies, individuals and politicians across Belgium to observe what is happening at the micro level. The tool was created using a combination of open datasets about mandates, addresses, company information and annual accounts.

## Environmental impact of open data

In a substantial part of the literature, environmental impact is not considered to be an independent category and is encompassed in the social and economic ones. Many studies (including those that refer to the ODB) place environmental impact under social impact, whereas only studies referring to the ODM use the environmental aspect independently to assess impact.

The European Commission's better regulation guidelines identifies environmental impact as changes in the state of the environment due to anthropogenic activities, including the use of resources or

<sup>(18)</sup> <https://vpt.lrv.lt/en/>.

<sup>(19)</sup> <https://vpt.lrv.lt/kovai-su-covid-19-sudarytos-sutartys>

<sup>(20)</sup> <https://www.stat.ee/en>

<sup>(21)</sup> <https://www.sm.ee/en>.

<sup>(22)</sup> <https://www.stat.ee/en/find-statistics/methodology-and-quality/esms-metadata/40701>.

<sup>(23)</sup> <https://www.mediafin.be/>

<sup>(24)</sup> <https://openthebox.be/>

activities causing pollution. Thus, this domain concerns water and air quality, noise levels in cities, waste management systems, environmental-friendly transport systems and carbon-dioxide reduction, all of which can be improved with the help of applications that use open data.

### Use cases that highlight environmental impact

One of the use cases that focus on environmental impact is the **German** application ‘Pegel-Online’ <sup>(25)</sup>, a service that raises awareness of water and air quality. This application is an open-data-based online platform that is offered by the German Federal Waterways and Shipping Administration <sup>(26)</sup> and provides information on the water level of more than 7 300 km of rivers and canals in Germany. Pegel-Online is a decision-making tool for citizens that live near water sources and for public agencies and officials that work on water-related issues in Germany.

Another example of open data that creates a positive environmental impact is the application ‘Kam s nim?’ <sup>(27)</sup> (where to put it?) in **Czechia**. ‘Kam s nim?’ is a non-profit project led by Czech association Let’s Clean <sup>(28)</sup>. It provides users with an interactive map that allows them to find locations where they can legally dispose of unwanted goods and waste, namely expired medicines, tires, batteries, lamps, bulk waste or hazardous waste.

### Spillover effects between impact domains

The previous subsections centred their attention on one impact domain for each use case. However, many spillover effects exist. It is rare that a use case will only have an impact on one domain; often, all four domains will be impacted one way or another. For example, several direct environmental impacts – such as climate impact, loss of biodiversity and the use of resources – have an (indirect) impact on the economy and wider social well-being. They impact economic activity through changes in production and services and impact social aspects through health, poverty and wealth distribution.

One example of such spillover effects is the PPO institution in **Lithuania**, which publishes data on public contracts regarding supplies and services to address the COVID-19 pandemic. Publishing data under an open licence contributes to the advancement of transparency and accountability, and therefore creates a political impact. Moreover, the PPO is contributing to economic development by sharing information about buyers, suppliers and prices within the market, enabling competitors to adjust accordingly. It also contributes to the social domain as it improves public health.

Another example is ‘Kam s nim?’, the application in **Czechia** that helps users to dispose of goods and waste in legal areas. Though this has been identified as impacting the environmental domain, it can be argued that the application can also have a social and economic impact. From a social perspective, citizens have access to clear information on where to dispose of waste, which will improve their quality of life. From an economic standpoint, businesses, including start-ups and SMEs, can create services around the information this application provides, such as alternative methods of waste disposal or alternative solutions to recycle or reuse the waste material.

---

<sup>(25)</sup> <http://pegelonline.wsv.de/gast/start>

<sup>(26)</sup> [https://www.gdws.wsv.bund.de/DE/startseite/startseite\\_node.html](https://www.gdws.wsv.bund.de/DE/startseite/startseite_node.html)

<sup>(27)</sup> <https://www.kamsnim.cz/>

<sup>(28)</sup> <https://www.data.gv.at/anwendungen/muell-checker/>

## 1.5. Conclusion: considerations for the development of an open data impact assessment methodology

Section 1 provided an overview of the articles and frameworks that define the impact or value created with open data. Most definitions of the impact of open data have a broad scope and cover a selection of the social, political, economic and environmental impact of open data. Moreover, important studies such as the ODM report have a specific focus on the impact that is created with open data.

Nevertheless, a common categorisation of impact domains and a common methodology to measure the impact of open data does not yet exist. The ODM covers social, political, economic and environmental impact, whereas the ODB, for instance, only focuses on social, political and economic impact.

In the absence of a common categorisation, we recommend structuring the impact assessment in line with the recommendations from the better regulation guidelines and toolbox <sup>(29)</sup>, which serve as the gold standard in the EU when preparing new initiatives. This means that the impact on government transparency or improvement in government efficiency – which would both be considered political impact in the ODM – should be classified as social impact and economic impact respectively. Using the better regulation guidelines, one can assess the impact of open data as follows.

- **Social impact of open data.** Improved public health, less corruption, more transparency and a higher quality of information.
- **Economic impact of open data.** Costs saved, jobs created and efficiency gains in the public sector.
- **Environmental impact of open data.** More sustainable energy use, improved biodiversity and a reduction of CO<sub>2</sub> emissions.

Besides the differences in the classification of impact domains, this section shows the different ways of scoping the measurement of impact. Some countries use a narrow scope for the impact of open data, mainly informed by web statistics such as the **number of applications** that have arisen from the use of the data, while other definitions also look at the **impact of the applications**. The recommendation in the development of an impact methodology is to not only measure the amount of use cases, but to focus on the impact created by the use cases.

---

<sup>(29)</sup> [https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how/better-regulation-guidelines-and-toolbox\\_en](https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how/better-regulation-guidelines-and-toolbox_en).

## 2. Impact measurement on national and municipal open data portals

To complement the literature review about open data impact as presented in Section 1, this one provides insights in the approaches implemented by leading European countries and municipalities to evaluate the impact of their open data portals. These portals were examined on the characteristics of three features enabling impact assessment: 1) reuse case sections, 2) statistics on data reuses and/or reusers and 3) indicators developed for impact studies <sup>(30)</sup>. The uncovered European **best practices** will serve as inspiration for the development of an impact assessment methodology for data.europa.eu.

At the national level, the open data portals of the following 12 countries are the subjects of analysis: Germany, Estonia, Ireland, Spain, France, Croatia, Cyprus, Lithuania, the Netherlands, Poland, Finland and Ukraine. The analysis of local-level practices equally concerns 12 portals: those of Barcelona, Berlin, Bordeaux, Dublin, Eindhoven, Florence, the Helsinki region, Lisbon, Paris, Vienna and Zaragoza. These 24 portals were selected following a quick scan of a wider range of portals identified on the basis of the 2021 Open Data Maturity Assessment results, open data research on cities and input from the open data community. The quick scan aimed to determine the presence of the three features of interest for the portal inventory: reuse case sections, use(r) statistics and impact studies. Annex I – Methodology for inventory of national and local open data portals provides further details on the research methodology, including the portal selection process and applied analytical framework. Annex II – Results of national portals

The remainder of this section is structured as follows. Section 2.1 presents the findings from national open data portals, while Section 2.2 describes those from the portal inventory at the local level. A more detailed breakdown of specific indicators for each portal can be found in Annex IV– Impact measurement on national and local open data portals. Drawing on the insights from the national and local portals, Section 2.3 concludes with considerations for the development of an open data impact assessment methodology.

---

<sup>(30)</sup> As a result of the structure of this report, these open data impact studies to which the portals refer overlap with the national open data impact studies mentioned in Section 1.

## 2.1. National open data portals

The data portals of the countries in Figure 1 will be discussed in this subsection.

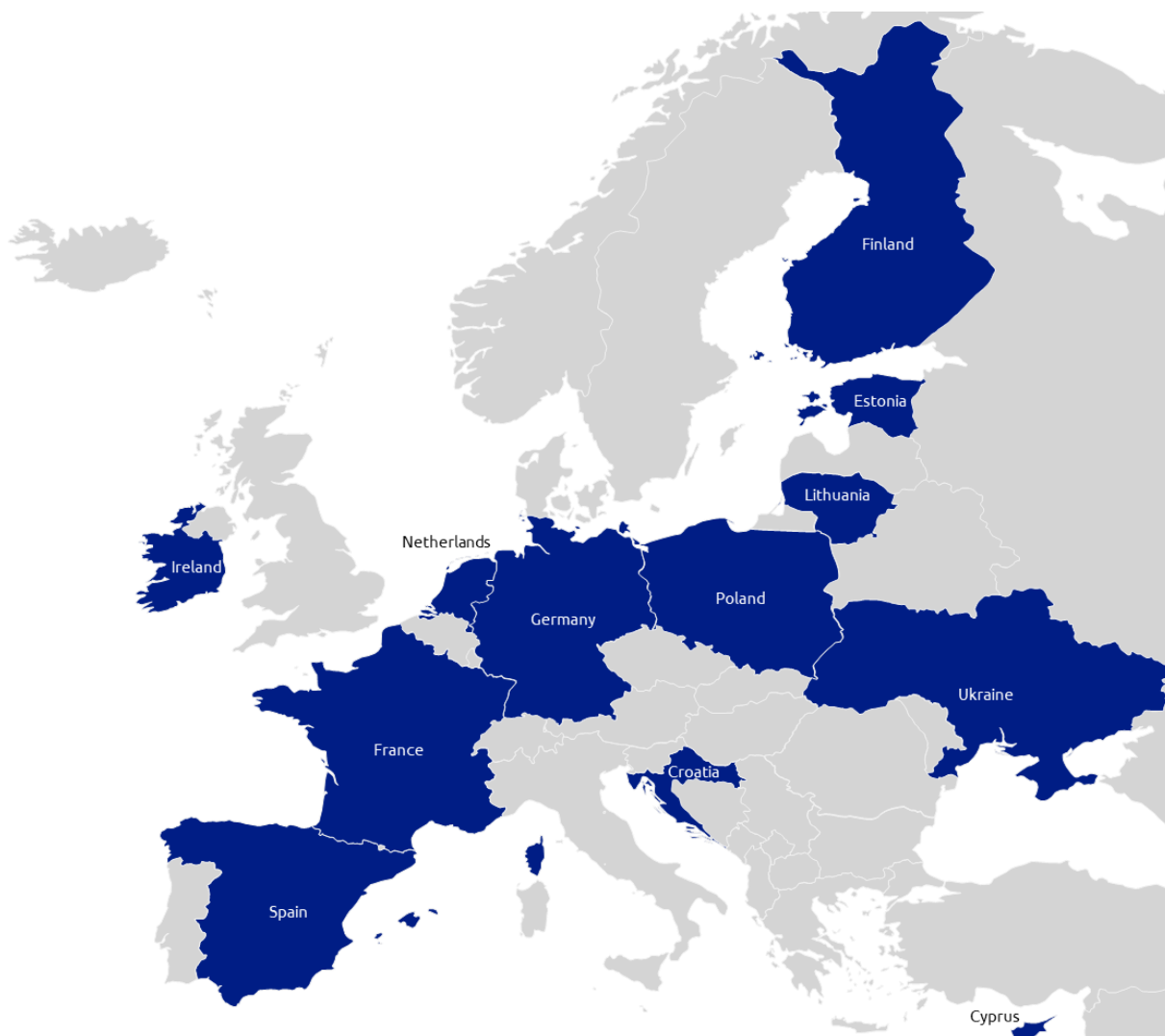


Figure 1. Overview of countries whose open data portals were included in this report

### Cyprus

The [Cypriot open data portal](#) enables the assessment of impact by compiling use cases and tracking portal statistics. Cyprus also conducted a formal survey-based impact study in 2021 <sup>(31)</sup>.

Under the 'Impact' tab, the portal catalogues 45 **reuse cases**, most of which are only available in Greek. Each case description includes the name of the developer, the type of reuse and the theme or themes under which the used dataset(s) fall. The portal provides no option to filter cases by type or theme. Figure 2 shows the description of Nomoplatform, an application that allows citizens to monitor parliamentary processes <sup>(32)</sup>. Because Nomoplatform works on multiple platforms, it has been tagged as a mobile app, a web app and a desktop app. Its theme, listed below the type, is *Κυβέρνηση και*

<sup>(31)</sup> Deloitte, 2021a, p. 50.

<sup>(32)</sup> Nomoplatform can be accessed at <https://www.nomoplatform.cy/> (in Greek). For a short review in English, see Charalambous (2021), 'NGO releases Nomoplatform™ electronic legislative observatory', *In-Cyprus* (<https://in-cyprus.philenews.com/news/local/ngo-releases-nomoplatform-electronic-legislative-observatory/>).

*Δημόσιος Τομέας*, meaning government and public sector. The use case section contains a link to a submission form for new cases (seen in Figure 2 above the case description), which does not provide criteria for inclusion beyond the incorporation of data from the portal. More detailed profiles of certain reuse cases are available on the portal's articles page, which may be accessed via a separate tab. Although these posts do not follow a consistent structure, they provide additional information on the selected cases. This may include links to the specific datasets used or application use statistics.

Figure 2. Reuse case section of the Cypriot national open data portal

Since May 2016, the portal's **statistics dashboard** has tracked the number of site visitors on a monthly basis and the number of listed reuse cases and requests for data biannually <sup>(33)</sup>. It also tracks the theme and source distribution of available datasets and provides a current ranking of datasets by number of user visits, although it is not clear how recent these figures are.

The portal provides links to **two impact-focused reports** published in 2021 <sup>(34)</sup>. The first is an impact study on open data in Cyprus <sup>(35)</sup>. The second is a handbook for an effective impact evaluation methodology based on that study's results <sup>(36)</sup>. The study aimed to evaluate the impact of open data on the social, political, economic and environmental sectors. Moreover, it assessed the nature and value of open data reuse within Cypriot organisations. It began with a literature review of publications on open data at both the national and European levels, including analytical reports from [data.europa.eu](https://data.europa.eu). Subsequently, a questionnaire was distributed to companies and organisations that made use of open data either internally or to provide services to the public. This group was identified through actor mapping during the preliminary research process. A total of 80 organisations responded during a 2-week period in February and March 2021. Respondents came from both the public and private sectors. The survey itself was composed of 32 questions falling under six categories: elements of the organisation/business, use of open data, employment/jobs, [capital] turnover,

<sup>(33)</sup> <https://www.data.gov.cy/node/4955?language=en>

<sup>(34)</sup> These reports are exclusively available in Greek ([https://www.data.gov.cy/Studies-Publications\\_EL?language=el](https://www.data.gov.cy/Studies-Publications_EL?language=el)).

<sup>(35)</sup> Deloitte, 2021a, p. 50.

<sup>(36)</sup> Deloitte, 2021b, p. 2.

benefits/advantages of using open data and social groups and environment. For the full text of questions assessing impact, see Annex IV– Impact measurement on national and local open data portals <sup>(37)</sup>.

The methodology handbook recommends four strategies for assessing the impact of open data on the national level. The first is the collection of primary data both from portal users and from surveys like the one carried out by the study (on a sample basis). The second is the ongoing collection and monitoring of portal use and user statistics. The third is the inclusion of reuse cases on the portal, and the fourth is regular communication with data providers and users to determine how the process of reuse might be streamlined and improved. The portal contains a general methodology section, which identifies four means through which the Cypriot government has evaluated and plans to continue to evaluate the impact of open data in the country. These are essentially identical to those laid out in the methodology handbook, although the portal page also recommends that studies akin to the two above-mentioned 2021 reports be regularly commissioned by national authorities, public data providers and municipalities.

## Germany

The [German open data portal](#) provides insight into impact by compiling reuse cases, which at the moment are available in German exclusively. Its **case repository** may be accessed through both a menu on the homepage and a widget displaying the total number of listed applications. Twenty-four recorded cases are labelled by application type (e.g. mobile app), theme (e.g. health) and operating system if applicable. Figure 3 shows the menu through which visitors may filter cases – note that Weigweiser Kommune, a guide to German municipalities, is correctly tagged as a website. The subpage for this case lists the eight themes under which it falls.

The portal has also implemented a more specific tagging system under which cases may be filtered by keywords, including ‘climate change’, ‘bicycles’ and ‘water levels.’ Once a visitor has filtered reuse cases by a keyword, they are provided with the option to further refine the selection by filtering by other tags attached to the results of the initial search. Thus, a visitor may choose to see all cases listed under ‘bicycle’ and then further filter by ‘theft’ to arrive at the case subpage for a website that publishes visualisations of data pertaining to bicycle theft in Berlin. Some inconsistencies are present that may complicate keyword-based filtering. Not all reuse cases tagged with ‘bicycle’, for example, are also tagged with ‘bike’, and vice versa. Only a fraction of the total number of keywords tagged can be viewed in the search options menu on the main reuse case page, which may hinder keyword-based filtering. Case subpages provide brief descriptions and in some cases links to the datasets used. As shown in Figure 3, a post on the main reuse case page solicits submissions for other applications reusing open data. Visitors may submit applications for inclusion through a contact form or by emailing the GovData coordination office. The portal provides no criteria regarding eligibility.

---

<sup>(37)</sup> Questions from the first category are omitted as they focus on organisational characteristics rather than impact.



The screenshot shows the German national open data portal interface. At the top, it displays '24 Treffer' (24 results) and navigation tabs for 'Alles' (53768), 'Daten' (53614), 'Anwendungen' (24), 'Informationen' (24), and 'Blog-Beiträge' (106). A search filter 'Erweiterte Suche' is active. On the left, there are filter options under 'Filtermöglichkeiten' with a 'zurücksetzen' button. The filters are categorized into 'TYP' (Type) and 'KATEGORIEN' (Categories). The 'TYP' filters include: Webseite (15), Visualisierung (11), Tool (7), Konzept (5), Mobile App (4), Sonstiges (2), and Wissenschaftliche Publikation (1). The 'KATEGORIEN' filters include: Bevölkerung und Gesellschaft (5), Bildung, Kultur und Sport (4), Gesundheit (2), and Landwirtschaft, Fischerei, Forstwirtschaft und Nahrungsmittel (2). The main content area shows two search results. The first result is titled 'Kennen Sie eine Anwendung, die hier noch fehlt?' (Do you know an application that is still missing here?) under the category 'Sonstiges'. It includes a call to action for users to report missing Open-Data applications. The second result is titled 'Wegweiser Kommune' (Municipality Navigator) under the category 'Webseite'. It describes a website providing statistics for all municipalities with more than 5,000 inhabitants as Open Data.

Figure 3. Reuse case section on the German national open data portal

Although the portal has no statistics dashboard, all keyword searches made between 2016 and 2019 were published as a dataset in 2021, allowing those who downloaded it to identify the most frequently searched terms over that period. However, the link to these data is no longer operational.

## Estonia

The [Estonian open data portal](#) provides insight into open data impact in various ways: through a use case section, a statistics dashboard and a collection of reports and articles.

The main **use case section** contains a visual overview of the submitted cases (see Figure 4). Visitors are provided with the option to filter cases by creator or by region, as well as to perform a general keyword search. They may click on each example to view its subpage, which contains a brief description and sometimes links to the dataset(s) used. These subpages also specify whether the applications update automatically over time as alterations are made to the datasets on which they are based, although most do not. Visitors may upload reuse cases to the portal through a form or by contacting an administrator by email.

Figure 4 Reuse case section on the Estonian national open data portal

The **use and user statistics dashboard** tracks open data publication through new data uploads by theme in the past 30 days, along with the percentage of all uploaded data falling under each theme <sup>(38)</sup>. This thematic breakdown is typical of national open data portals and includes divisions such as ‘Science and technology’ and ‘Health’. Regarding open data impact, a ranking of all available datasets by number of user downloads is also provided. This page similarly contains data from the past 30 days and seems to update regularly. A menu on this dashboard links to a Google Analytics subpage on which visitors may view the number of unique users, the number of unique searches, the most common cities from which site visitors originate and a ranking of the 100 most searched-for keywords within the portal over the last 28 days.

Although the portal lacks a defined impacts section, a variety of reports and articles on the impact of open data in Estonia may be found <sup>(39)</sup>. In November 2020 and December 2021, the Estonian Ministry of Economic Affairs and Communications invited 64 and 94 public sector agencies respectively to respond to a survey on their approach to open data. The seven open-ended questions asked the agencies concerned about their ability to assess the nature of data reuse and the benefits of publishing open data for agencies and users. For the full text of these questions, see Annex IV. Due to the difficulty of evaluating the impact of open data quantitatively (e.g. through indicators such as businesses’ financial savings), the Ministry chose to focus on **mapping benefits through stories** <sup>(40)</sup>. It is worth noting that only 44 % and 16 % of selected agencies responded to the survey in 2020 and 2021, respectively <sup>(41)</sup>. It is unclear whether the survey will take place on a yearly basis.

Visitors may also access two articles on specific applications of open data: enabling machine translation of texts into Estonian and transitioning towards a sustainable economy. The Ministry provides

<sup>(38)</sup> <https://avaandmed.eesti.ee/statistics>

<sup>(39)</sup> See <https://avaandmed.eesti.ee/instructions> (in Estonian). The English language version of this page does not display all articles shown on the Estonian-language page, notably the mentioned impact assessment reports.

<sup>(40)</sup> Ministry of Economic Affairs and Communications, 2021.

<sup>(41)</sup> Ministry of Economic Affairs and Communications, 2021.

guidelines for assessing the impact of open data in a post in the same section. It suggests measuring the number and frequency of downloads, identifying common subjects of information requests made to organisations and evaluating user satisfaction. User reviews of the portal may also be viewed on the statistics dashboard, although only four visitors have reviewed the portal since its inception. Logged-in users have the option to rank the portal's quality on a 10-point scale and to provide additional written feedback.

## Spain

The [Spanish open data portal](#) has an impact section divided into four sub-sections: reuse companies, reuse applications, an open data initiative map and a statistics dashboard.

The **reuse companies** subsection lists 80 companies that make use of data from the portal<sup>(42)</sup>. They are labelled by category (e.g. transportation) and by the amount of time they have been established (e.g. more than 20 years). A list of over 70 divergent tags, referring to either the nature of the service (e.g. 'web apps') or its area of focus (e.g. 'tourism' or 'fuel price'), provides another way of filtering the reuse companies.

The **applications** subsection lists 398 reuse cases, which can be filtered according to the same categories and tags, and also specifies the publication date and platform or operating system of each case (see Figure 5)<sup>(43)</sup>. Once a visitor has filtered reuse cases by a keyword, they are provided with the option to further narrow down the selection by filtering by other tags attached to the results of the initial search. Thus, a visitor may choose to see all cases listed under 'economy' and then further filter by 'treasury' to arrive at the case subpage for an online platform that provides access to data on public contracts offered within the European Union. Some inconsistencies are present that may complicate keyword-based filtering. For example, 'covid19' and 'covid-19' are tagged separately rather than consolidated under a single label. Case subpages consistently contain links to source datasets. The initiative map subsection allows visitors to view a map of the country marked with the locations of 309 open data initiatives run by public administrations. These are tagged by subject and by administration level (e.g. local or regional). Many of these initiatives concern the supply side of open data, such as the institutional open data strategy. Others also refer to the demand side, pointing for example to the applications catalogue of the institutional open data portal.

---

<sup>(42)</sup> <https://datos.gob.es/en/casos-exito>

<sup>(43)</sup> <https://datos.gob.es/en/aplicaciones>

## Applications



This section contains some examples of applications and solutions developed using, among others, data supplied by the public administrations. The Aporta Initiative does not guarantee the accuracy of the information contained in each description page or the website that describes the application in more detail.

The screenshot shows the 'Applications' section of the Spanish national open data portal. On the left, there is a 'Category' list with the following items and counts: Transport (69), Public sector (62), Environment (51), Culture and leisure (41), Tourism (37), Economy (28), Town planning and infrastructures (28), Rural environment (27), and Society and welfare (27). The main content area displays a search bar with a 'SEARCH' button. Below the search bar, it indicates '406 applications found' and 'Sort by: Date descending'. A search result for 'Visor Municipal de Sagunto' is shown, categorized under 'Ayuntamiento de Sagunto' and 'Public sector'. The description states: 'El Visor Municipal de Sagunto es una herramienta web que permite consultar de una manera gráfica e interactiva los presupuestos del municipio. Gracias a esta aplicación, se facilita a la ciudadanía la visualización y comprensión de las cuentas del...'. A small thumbnail image of the application is also visible.

Figure 5. Reuse case section on the Spanish national open data portal

The **initiative map** allows visitors to view a map of the country marked with the locations of 309 open data initiatives run by public administrations <sup>(44)</sup>. These are tagged by subject and administration level (e.g. local or regional). Many of these initiatives concern the supply side of open data, such as institutional open data strategies. Others also refer to the demand side, linking for example to applications catalogues on institutional open data portals.

The **statistics dashboard** shows the monthly change in the number of cases recorded on the portal since March 2013 and in the number of initiatives and reusing companies recorded since January 2017 <sup>(45)</sup>. The portal also tracks change in data uploads by theme and by publisher (e.g. state administration or private entity), as well as the 10 most frequently visited datasets; these figures are cumulative since 1 December 2016. Visitors may also view the 10 most frequently visited datasets in any specific month from December 2016 to present day. All metrics are updated daily.

Users may add examples of reusing companies, applications or public initiatives via separate, quite detailed forms available via the interact/report section. Selected reuse cases are promoted through the 'News' section, a quarterly newsletter and the data portal's blog. The Spanish portal is particularly good at explaining the factors that make the selected cases particularly impactful or exemplary, for instance as regards EU policy goals.

## Finland

The **Finnish open data portal** compiles reuse case examples and tracks portal statistics on a designated dashboard.

The **'Showcases'** section includes 83 cases, which are briefly described and labelled by theme (e.g. health) and operating environment, if applicable <sup>(46)</sup>. Figure 6 shows the menu through which visitors may filter cases thematically. This section also allows filtering through a more general keyword-based tagging system; keywords can describe either the nature of the service (e.g. 'visualisation') or its subject matter (e.g. 'cars'). The keyword system seems to have been implemented with relative

<sup>(44)</sup> <https://datos.gob.es/en/iniciativas>

<sup>(45)</sup> <https://datos.gob.es/en/dashboard>

<sup>(46)</sup> [https://www.avoindata.fi/data/en\\_GB/showcase](https://www.avoindata.fi/data/en_GB/showcase)

consistency – applications dealing with train and bus services, for example, are consistently tagged with ‘public transport’. That said, some discrepancies are present – a real-time map created by Traffic Management Finland using portal data lacks the ‘trains’ tag, despite providing railway traffic updates, while other traffic-related applications are consistently tagged with the type of vehicle that they monitor. The subpage for each reuse case links to the dataset(s) used. Visitors wishing to submit new applications for inclusion must fill out a form requiring them to provide the application’s purpose and process of development, along with the ways in which it uses open data. Figure 6 shows a request for new submissions at the top right.

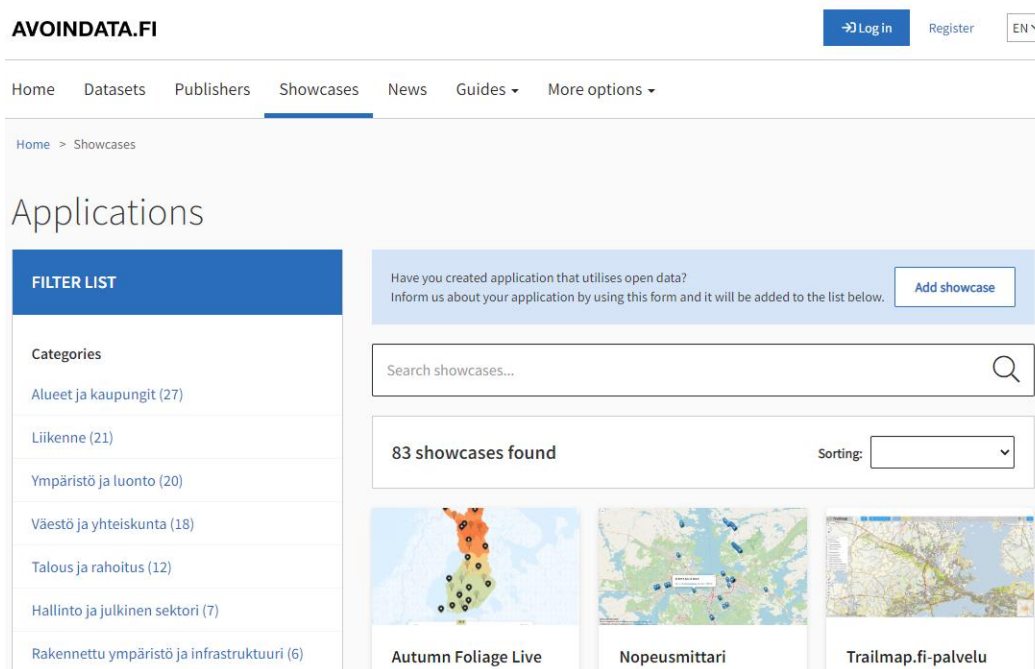


Figure 6. Reuse case section on the Finnish national open data portal

The portal tracks **statistics** in eight categories: 1) administrative branch summary, 2) audience locations, 3) outdated datasets, 4) least popular datasets, 5) most popular datasets, 6) most popular organisations, 7) most popular resources and 8) most popular search terms<sup>(47)</sup>. Visitors may view metrics from the last week, month or year. All metrics are updated daily.

## France

The [French open data portal](#) contains a reuse case section and tracks a wide range of portal statistics, allowing for the assessment of impact.

The **reuse repository** features 3 204 cases categorised by theme (e.g. health) and type (e.g. visualisation)<sup>(48)</sup>. Despite this dual system of categorisation, as shown in Figure 7, the general reuse case page only allows visitors to filter cases by theme. Case subpages contain brief descriptions of the applications along with links to the datasets used. Visitors may leave feedback concerning these applications on the subpages through text submission boxes and may also reply to questions and comments left by previous visitors. The portal provides comprehensive guidelines for publishing new reuse cases. Logged-in users can submit applications for inclusion either independently or on behalf of

<sup>(47)</sup> [https://www.avoindata.fi/data/en\\_GB/report](https://www.avoindata.fi/data/en_GB/report)

<sup>(48)</sup> <https://www.data.gouv.fr/en/reuses/>

an organisation <sup>(49)</sup>. In addition to the necessary information for categorisation, the submission form requests descriptions of the process of development, the conclusions of the project and the developer or developing organisation. Here users are also asked to provide feedback on the process of reuse and to explain the ways in which available data could be improved to enable further use or analysis.

The screenshot shows the 'Reuses' section of the data.gouv.fr portal. At the top, there is a search bar and navigation links for 'Data', 'Reuses', 'Organizations', 'Getting started on data.gouv.fr', 'News', and 'Contact us'. Below the search bar, there are several category buttons: 'All', 'Health', 'Transport and mobility', 'Housing and development', 'Food and agriculture', 'Culture and recreation', 'Economy and business', 'Environment and energy', 'Work and training', 'Politics and public life', 'Safety and security', 'Education and research', 'Society and demography', 'Law and justice', 'Open data tools', and 'Others'. The page indicates '3205 results' and a 'Sort by: Relevance' dropdown. Three preview cards are shown below: a map of France, a detailed map of a city area, and a card for 'L'Annuaire des Entreprises' (Business Directory) with a search bar.

Figure 7. Reuse case section on the French national open data portal

A **statistics page** using the Matomo analytics platform exists, but is not accessible from the portal homepage <sup>(50)</sup>. Statistics are divided into three categories, each of which may be viewed on a separate subpage: visitors, behaviour and acquisition. The 'Visitors' subsection tracks visits to the site over time. Viewers can set custom date ranges from within which to view data. The total numbers of pageviews, unique pageviews, searches and unique keyword searches are also provided. The four pages that have experienced the most dramatic changes in number of views are ranked in a separate list. This page contains a wide array of visitor-related statistics, primarily of a geographic (e.g. country of origin) or technical (e.g. device type and model) nature. The 'Behaviour' subsection tracks page visits and downloads by URL and provides further statistics on the former (e.g. average visit duration). The 'Acquisition' subsection tracks the means through which visitors enter the site (e.g. directly or from search engines). The portal's administrators publish yearly summaries of changes and accomplishments in posts that may be found in the 'News' section of the portal. These do not always make reference to portal data, but some, like the 2019 summary, have pointed to significant increases in site visitors and data downloads as evidence of impact.

<sup>(49)</sup> <https://guides.etalab.gouv.fr/reutilisation/publier-reutilisation/#comment-referencer-une-reutilisation>

<sup>(50)</sup> [https://stats.data.gouv.fr/index.php?module=CoreHome&action=index&idSite=109&period=range&date=previous30#?idSite=109&period=range&date=previous30&segment=&category=Dashboard\\_Dashboard&subcategory=1](https://stats.data.gouv.fr/index.php?module=CoreHome&action=index&idSite=109&period=range&date=previous30#?idSite=109&period=range&date=previous30&segment=&category=Dashboard_Dashboard&subcategory=1)

## Croatia

The [Croatian open data portal](#) provides two mechanisms for the assessment of impact: a reuse case repository and a use and user statistics dashboard.

The **data reuses** menu points to three subpages: best practice examples, applications and reuse examples. Two best-practice examples are mentioned: a Slovenian competition to create applications that promoted better integration of cultural heritage into tourist activities and a Croatian cloud-based application for managing support processes for kindergartens <sup>(51)</sup>. The six showcased applications (see Figure 8) are labelled by type (i.e. application) and by whether they are free to use <sup>(52)</sup>. The subpage for each contains a brief description of the functions of the application. The reuse examples page is still empty, but visitors may submit new reuse cases for inclusion through a form.

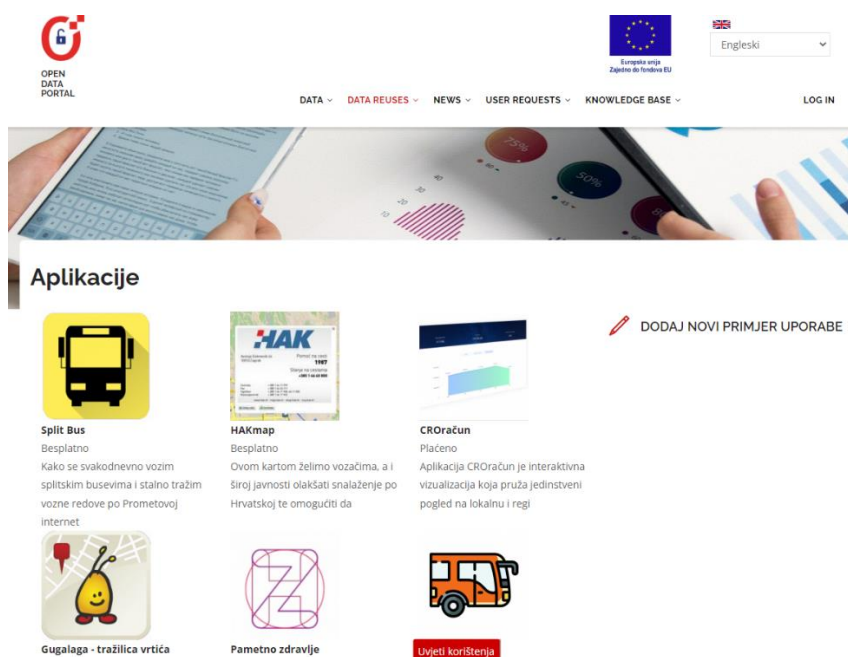


Figure 8. Reuse case section on the Croatian national open data portal

The portal's use and user **statistics dashboard** tracks the number of publishers, datasets, registered users, and visits to the site over time <sup>(53)</sup>. The number of datasets by topic is also displayed in pie chart format, with topics including 'Economy and finance' and 'Energy.' The number of usage examples by type is also tracked, with the first and last update dating from December 2021. The portal contains pages that list and solicit public suggestions for publication of new datasets and correction of or upgrades to existing datasets. The number of suggestions for new datasets and the number of suggested corrections are tracked graphically on the statistics dashboard, but on separate subpages under the 'Portal data' heading. Visitors may view the overall success in resolving these suggestions by dataset topic and dataset publisher, expressed in percentage form. Only three suggestions appear to have been made thus far, with two new datasets accepted and one rejected. Consequently, these metrics are based on very limited data. Statistics appear to be updated at least on a monthly basis; all

<sup>(51)</sup> <https://data.gov.hr/en/primjeri-dobre-prakse>

<sup>(52)</sup> <https://data.gov.hr/en/aplikacije>

<sup>(53)</sup> <https://data.gov.hr/en/ga-statistics-report>

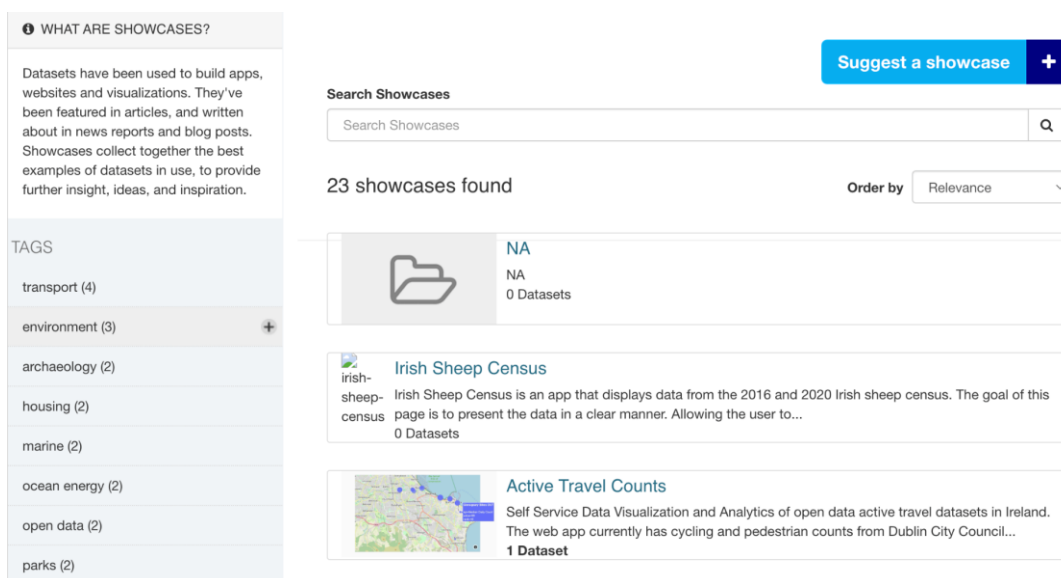
are up-to-date, although some graphs erroneously label the month following June 2022 as December 2021.

## Ireland

The [Irish open data portal](#) allows impact to be assessed in two ways: through a reuse case repository and through portal statistics.

The portal’s **showcases section** displays 24 cases categorised by means of a keyword-based tagging system shown in Figure 9 <sup>(54)</sup>. Keywords may describe either the nature of the service (e.g. ‘android app’) or its subject matter (i.e. ‘agriculture’). Once a visitor has filtered reuse cases by a keyword, they are provided with the option to further narrow down the selection by filtering by other tags attached to results of the initial search. Thus, a visitor may choose to see all cases listed under ‘transport’ and then further filter by ‘Irish Rail’ to arrive at the case subpage for a webpage that shows the live position of trains across Ireland.

Since keywords are suggested by the submitter when a new case is submitted for inclusion and not standardised afterwards, there are some **inconsistencies in tagging** that may complicate keyword-based filtering. For example, ‘Capital Acquisitions Tax’ and ‘capital acquisition tax’ are tagged separately rather than consolidated under a single label. A total of 11 case subpages contain links to the dataset or datasets used. The amount of information provided on these subpages varies widely – the form through which visitors may submit new reuse cases for inclusion encourages detailed description of the case, but provides no guidance as to the information that should be included. Consequently, some pages, like that of the application Failte Maps, detail the case’s creator, function, purpose, process of development and technical specifications, while others, like that of the web application Active Travel Counts, only briefly summarise its nature. Two cases, a dashboard for data on hospital waiting times and a visualisation of DublinBikes usage over the course of each day, have been highlighted under a separate ‘Impact Stories’ section, but no further detail is provided regarding their nature, impact or development process. These cases are not included in the general reuse case repository.



<sup>(54)</sup> <https://data.gov.ie/showcase>



Figure 9. Reuse case section on the Irish national open data portal

The portal's use and user **statistics dashboard** tracks changes in the number of users and number of datasets added from 2016 to present day <sup>(55)</sup>. Visitors may also consult the 10 most-viewed datasets, most-downloaded datasets and most frequently searched-for keywords, although it is not clear how recent this information. Datasets suggested but not yet approved for inclusion on the portal are listed elsewhere under a 'Suggested datasets' section. The statistics dashboard has tracked the number of new and resolved suggestions on a bimonthly basis since August 2016, but this metric has not been updated since September 2021, despite the most recent suggestion listed dating to May 2022. All other dated metrics appear to have last been updated in April 2022. Monthly website statistics from January to May 2022 may be viewed or downloaded in CSV format; these metrics include the number of visitors, number of sessions and average session duration.

## Lithuania

The [Lithuanian open data portal](#)'s impact may be tracked in two ways: by means of its reuse case repository and through a set of compiled statistics published in 2021.

The portal contains a **reuse case** section profiling 28 cases, which are divided into 'Use case examples use' and 'Created applications', although no cases are currently listed on the subpage for the latter <sup>(56)</sup>. An explanation of the service or services provided by each case is supplied, as are a brief description of its benefits and a list of beneficiaries.

In Figure 10, the benefits of a systematic model for forecasting regional labour market needs created by Neurotechnology are listed under *Teikiama nauda*; its beneficiaries (the city of Panevėžys, its municipal administration and its citizens) are listed under *Naudos gavėjų grupė*. In some cases, the general description details further benefits: in this case, the website's contributions to the development of a broader 'citizen-centred public service strategy' in Lithuania. Links to the datasets used are provided for two of the 28 cases. There is also an 'additional information' section which supplies information about the data used when links to specific datasets are not provided (e.g. it is specified that a website that provides real-estate consultations uses 'data collected from municipalities, state institutions, heat supply companies and other organisations'), implying that not all listed use cases are based on open data available on the portal. Additional information is provided for a further nine cases. Logged-in visitors may submit new reuse cases through a form.

<sup>(55)</sup> <https://data.gov.ie/stats>

<sup>(56)</sup> <https://data.gov.lt/usecases/examples>

DATA.GOV.LT

Datasets Use cases Requests and suggestions More Log in LT

Home → Panaudojimo atvejai

## Use case examples

Use cases

- Use case examples
- Created applications

Sharing best practices on open data adds value and can inspire other data users to follow your example and create new services or otherwise increase the value of open data reuse.

Submit a new use case

<https://www.neurotechnology.com/>

**Description of use case/service:**

Vienas iš pagrindinių iššūkių, su kuriais šiandien susiduria Panevėžio miestas ir regionas – kvalifikuotos darbo jėgos, atitinkančios darbo rinkos poreikius, stoka. 2020 m. Panevėžio plėtros agentūros užsakymu atlikta studentų skaičiaus didinimo Panevėžio mieste studija atskleidė, kad aukštosiose ir profesinėse mokyklose nesurenkamos studentų / mokinių grupės į aktualias regionui programas neleidžia efektyviai išnaudoti švietimo potencialo mieste, o šiai problemai spręsti būtina pradėti sistemiskai rinkti duomenis, juos analizuoti ir stebėti bei jais remiantis įgyvendinti strategines regiono iniciatyvas. Tam buvo ieškoma sistemizuoto regiono darbo rinkos poreikių prognozės modelio. Įmonės "Neurotechnology" sukurtas sprendimas leidžia automatinio būdu rinkti laisvai prieinamus duomenis ir daryti išvadas apie darbo rinkos situaciją ir ateities poreikius. Darbo rinka yra kompleksinė socioekonominė sistema, todėl yra svarbu tinkamai pasirinkti labiausiai reprezentatyvius duomenų šaltinius ir taip pat parinkti tinkamus giliojo mokymosi algoritmus tiksloms ateities prognozėms.

**Benefits:**

Sprendimas visapusiškai ir sistemingai prognozuoja regiono darbo rinkos poreikius, leidžia tikslingai organizuoti vietos profesinio orientavimo paslaugas bei formuluoti mokymo ir darbuotojų perkvalifikavimo paslaugas. Naudojantis sukurtu sprendimu reikalaujama mažiau žmoniškųjų ir finansinių resursų.

**Beneficiary group:**

Panevėžio miestas, Panevėžio miesto savivaldybės administracija, piliečiai

**Additional information:**

**Used datasets:**

Likes: 0

Figure 10. Reuse case section on the Lithuanian national open data portal

Although the portal has no statistics dashboard, **use statistics** have been compiled and published as part of a 2021 review on open data publication conducted by the Information Society's Development Committee<sup>(57)</sup>. This report tracks changes in the number of portal users, institutional coordinators and data processors and in the number of registered datasets, open datasets and datasets with metadata available. For each of the 14 defined themes (e.g. economy and finance) the number of new datasets added to the portal and downloads in 2021 are also provided. The annexes of the report contain a table showing the number of new uploads to the portal by independent data providers. While open data impact is not the main focus of the review, the news item presenting some highlights of the report mentions a concrete estimation of economic open data impact without providing a foundation for this statement: 'It is estimated that the rational use of data can additionally create about 2 percent of national GDP'<sup>(58)</sup>. Lithuania has also announced the upcoming publication of a report on the progress and impact assessment of data openness<sup>(59)</sup>. Details on the methodology and results of the survey sent to public sector institutions and Lithuanian municipalities in the spring of 2022 are not yet available at the time of writing this report.

## Netherlands

A summary of the ways in which the [Dutch open data portal](#) allows the impact of open data to be assessed is available on the portal's impact section. It lays out four means of obtaining insight into the impact of open data on a national level: through data applications, impact stories, statistics and data communities<sup>(60)</sup>.

<sup>(57)</sup> Belickas, 2022

<sup>(58)</sup> <https://data.gov.lt/news/25> (in Lithuanian)

<sup>(59)</sup> <https://data.gov.lt/news/27> (in Lithuanian)

<sup>(60)</sup> Only available in Dutch at the time of writing this report.

The portal contains a **reuse case section** listing 118 applications <sup>(61)</sup>. These may be filtered by the data reuser, indicated as ‘application data owner’ (e.g. the Central Bureau of Statistics), by theme (e.g. traffic) or most recent update (e.g. last month, last year or longer than one year ago). Cases are categorised under 14 general themes, 10 of which have between one and four subthemes; within the ‘Housing’ theme, for example, visitors may further refine their search by ‘Buy and sell’ or ‘Renting and letting’. Figure 11 shows a portion of the theme menu on the left. Each case subpage contains a brief description and lists the theme, application type (e.g. website) and reuser type (e.g. a government organisation or a market party). When a case is categorised under a subtheme, the overarching theme is not listed on its subpage. Subpages also consistently link to the dataset or datasets used. Visitors may submit new reuse cases for inclusion via a form.



Figure 11. Reuse case section on the Dutch national open data portal

The portal’s **statistical dashboard** tracks the total number of published datasets from January 2021 to June 2022 <sup>(62)</sup>. Datasets are also displayed by provider in pie chart format. A subpage on the dashboard displays rankings of the 10 most-viewed datasets every year since 2016. Visitors may also view the 10 most frequently searched-for topics in 2021, 2020 and 2019.

Furthermore, in line with the data.europa 2021 ODM report <sup>(63)</sup>, the Dutch framework outlines the ways in which impact can be monitored in four areas: the social, political, economic and environmental areas. **Impact stories** falling within all four areas are available on the portal <sup>(64)</sup>. These stories take the form of interviews with individuals who have either created applications that reuse open data or worked with organisations that have done the same. The exact wording of the questions asked varies, but generally these interviews aim to lay out the function and purpose of the reuse application, the ways in which it makes use of open data and its practical benefits to users. Interviewees are also asked to discuss challenges they encountered during the process of data access and reuse and to suggest ways in which the portal could better serve the needs of reusers. For example, in an interview with its creator in 2021, [moethetraamdicht.nl](https://data.overheid.nl) (shouldthewindowbeclosed.nl) was highlighted as an example of an initiative with social impact <sup>(65)</sup>. This website, which informs inhabitants of Schiphol about expected air traffic at night, makes use of flight, noise mapping and resident complaint datasets. Its

<sup>(61)</sup> <https://data.overheid.nl/community/toepassingen> (in Dutch)

<sup>(62)</sup> <https://data.overheid.nl/statistieken> (in Dutch)

<sup>(63)</sup> Publications Office of the European Union, 2022a.

<sup>(64)</sup> <https://data.overheid.nl/assessment-impact-data-2021>

<sup>(65)</sup> <https://data.overheid.nl/actueel/impact-story/impact-story-moet-het-raam-dicht> (in Dutch)

creator, Sjoerd van den Hoorn, notes that incomplete data complicated the reuse process, but that the existence of the portal facilitated his search for relevant datasets.

In 2021, a **'data communities'** section was added to the portal. The function of these communities is to consolidate datasets, data requests, reuse cases and information on reusers and publishers related to a specific theme. Five communities are currently available or under development: energy, mobility, social security, education and migration. Updates and news on each subject may be found on its respective community subpage. Communities also provide a space for public discussion of relevant data and facilitate interactions between publishers, reusers and subject experts.

When a visitor searches for data in the portal, datasets with attached reuse cases or impact stories will be prioritised in results sorted by 'relevance', which is the default setting. Several years before the mentioning of high-value datasets in the open data directive<sup>(66)</sup>, the Dutch portal already started identifying and tagging certain datasets as 'high value'. The Dutch interpretation of "high value" is the degree to which datasets contribute to transparency, support a legal obligation, help reduce costs, may benefit a specific target audience or have a potential for reuse<sup>(67)</sup>. Municipalities and provinces maintain independent high-value data lists. Additionally, portal administrators have inventoried all data published on the portal to identify high-value datasets not included in these lists. The criteria for this classification are inspired by various international benchmarks, including the G8 Open Data Charter, which lays out 14 high-value data categories and provides examples of high-value datasets for each one<sup>(68)</sup>. For example, under the broader theme of 'Social mobility and welfare', data related to unemployment benefits is considered as high value.

## Poland

The [Polish open data portal](#)'s primary mechanism for assessing impact is its **PoCoTo data reuse** section. The 53 profiled cases are briefly described and labelled by type (website, application or other)<sup>(69)</sup>. As shown in Figure 12, 28 applications and 22 websites are currently listed. Of the three cases falling under 'other', one is a plug-in for a database of Polish waters, one is an energy and mining data platform and the third is a municipal open data portal. Information on the reuser and licence type is included on the case subpages. The service has also implemented a more **specific keyword-based tagging system**, where keywords may describe either the nature of the service (e.g. 'search engine') or its subject matter (e.g. 'kindergartens' or 'credibility of companies'). Since keywords are suggested by the submitter when a new case is proposed for inclusion, there are some inconsistencies in tagging that may complicate keyword-based filtering. For example, two similar applications that provide the real-time locations of trams and buses are tagged with 'public transport', but only one has the more general 'transport' tag. Additionally, four cases have no keyword tags at all. Although links to specific datasets used are not provided, some case descriptions provide more general information on data sources. The description provided for [energy.instrat.pl](#), a platform that compiles and represents energy- and mining-related data, for example, specifies that the figures provided are taken from 'public statistics, publicly available databases aggregating many public sources... and stock exchange data' and provides specific sources for each. New reuse cases may be submitted for inclusion through a form. This form provides an option to specify the datasets used, either by searching within the portal or

<sup>(66)</sup> <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019L1024>

<sup>(67)</sup> <https://data.overheid.nl/en/community/maatschappij/high-value>

<sup>(68)</sup> G8, 2016.

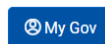
<sup>(69)</sup> <https://dane.gov.pl/en/showcase>

linking to an external source of open data, so this information can be included on case subpages. A number of further **reuse cases developed through datathons** sponsored by the Chancellery of the Prime Minister are profiled separately in articles available in the portal's 'News' section. Since these are prototypes rather than developed applications, they are not cross-listed in the data reuse section.

The [Polish open data portal](#)'s primary mechanism for assessing impact is its **PoCoTo data reuse** section. The 53 profiled cases are briefly described and labelled by type (website, application or other) <sup>(70)</sup>. As shown in Figure 12, 28 applications and 22 websites are currently listed. Of the three cases falling under 'other', one is a plug-in for a database of Polish waters, one is an energy and mining data platform and the third is a municipal open data portal. Information on the reuser and licence type is included on the case subpages. The service has also implemented a more **specific keyword-based tagging system**, where keywords may describe either the nature of the service (e.g. 'search engine') or its subject matter (e.g. 'kindergartens' or 'credibility of companies'). Since keywords are suggested by the submitter when a new case is proposed for inclusion, there are some inconsistencies in tagging that may complicate keyword-based filtering. For example, two similar applications that provide the real-time locations of trams and buses are tagged with 'public transport', but only one has the more general 'transport' tag. Additionally, four cases have no keyword tags at all. Although links to specific datasets used are not provided, some case descriptions provide more general information on data sources. The description provided for [energy.instrat.pl](#), a platform that compiles and represents energy- and mining-related data, for example, specifies that the figures provided are taken from 'public statistics, publicly available databases aggregating many public sources... and stock exchange data' and provides specific sources for each. New reuse cases may be submitted for inclusion through a form. This form provides an option to specify the datasets used, either by searching within the portal or linking to an external source of open data, so this information can be included on case subpages. A number of further **reuse cases developed through datathons** sponsored by the Chancellery of the Prime Minister are profiled separately in articles available in the portal's 'News' section. Since these are prototypes rather than developed applications, they are not cross-listed in the data reuse section.

---

<sup>(70)</sup> <https://dane.gov.pl/en/showcase>



#I helpUkraine

FIND OUT MORE

AND AND AND AND

PL EN | Register / Log in





Homepage Date Publishers PoCoTo News Knowledge base

You are here: [home](#) / PoCoTo

## Date reuse

If you have created or you know an application, website or other product that uses data, among others from the [Dane.gov.pl](#) portal and you want to share information about them on the portal - click the [Suggest](#) button and fill in the form. We will post your proposal on the [Dane.gov.pl](#).

Below we present examples of public data utilization in applications, products and services. We do not promote a specific solution, we only disseminate information about selected projects using data.

Search ...  

**Filters**

**Category**

Search ...


Application (28)


Other (3)

Website (22)

Apply


**Found:** Data reuse (53)

**Sort by:** Availability date 



[changeakotla.pl](#)

The website provides 3 e-services: eKOGRANT: e-service facilitating the submission of an application for co-financing for the replacement of the heating source to the Clean Air program and to the municipal project. Provides replacement

 Website

**Availability date:**  
17 August 2022, 11:13




Figure 12. Reuse case section on the Polish national open data portal

## Ukraine

Insight into open data impact in Ukraine is provided through a designated impact section and news articles on the Ministry of Digital Information's broader open-data-related [website](#) and to a lesser extent through portal statistics presented on the country's [open data portal](#) <sup>(71)</sup>.

The **designated impact section** evaluates the impact of open data in 11 areas: construction, infrastructure, health, ecology, business, the legal branch, local government, state supervision, financial transparency, forestry and the property and income declarations of public officials. The subpage for each area both broadly discusses the ways in which open data may be applied and provides links to relevant reuse applications and visualisations. On the infrastructure subpage (see Figure 13), for example, links are provided to an application that provides public access to infrastructure planning, a portal for monitoring road construction expenditure and a map that allows users to check the status of repair work across Ukrainian localities. The functions of these applications and their potential benefits are detailed. In some cases, realised impact is also discussed. The local government subpage, for example, notes that analysing electronic ticket data from Zhytomyr has allowed officials to save more than the equivalent of EUR 16 000 euro (UAH 500 000) per month through more efficient public transit routing <sup>(72)</sup>. The subpage for the legal branch describes two high-profile corruption cases that were brought to light with the aid of open data and provides links to relevant articles and news

<sup>(71)</sup> Even though Ukraine's open data portal and website are not operational at the time of writing this report, the country's activities on open data impact are included in the analysis for this report by using archived versions available on <https://web.archive.org/>. By highlighting Ukraine's open data efforts, data.europa.eu stands with Ukraine ([https://eu-solidarity-ukraine.ec.europa.eu/eu-stands-ukraine\\_en](https://eu-solidarity-ukraine.ec.europa.eu/eu-stands-ukraine_en)).

<sup>(72)</sup> Based on the exchange rate on 1 January 2022, see [https://ec.europa.eu/info/funding-tenders/procedures-guidelines-tenders/information-contractors-and-beneficiaries/exchange-rate-infoeuro\\_en](https://ec.europa.eu/info/funding-tenders/procedures-guidelines-tenders/information-contractors-and-beneficiaries/exchange-rate-infoeuro_en).

broadcasts. The two to five most popular datasets pertaining to each area are also listed. In addition to the impact section, **specific reuse cases are profiled** in articles published under a 'News' section on the [diia.data.gov.ua](https://diia.data.gov.ua) homepage. The formatting of these profiles varies, but most describe the nature of the service and its potential benefits. Some also specify intended beneficiaries.

## Сервіси та додатки



Figure 13. Reuse applications on the infrastructure impact subpage of the Ukrainian open data portal

As part of the transparency and accountability in public administration and services programme supported by the United States Agency for International Development and by aid from the UK government, the Ukrainian Ministry of Digital Transformation has published **nine reports on the impact of open data** in the areas of state supervision, financial transparency, forestry, construction, roads, healthcare, ecology, business and local government. These studies' conclusions were based on interviews conducted with the managers of reuse services, government agency representatives and other relevant actors in early 2020. The state supervision study, for example, interviewed 17 'key respondents' on the application of data in these areas: eight from the private sector (e.g. application developers) and nine from public administrations (e.g. department heads and inspectors) <sup>(73)</sup>. While transcripts are not provided and interview format seems to have varied, quotes from these reuse application creators and managers regarding the nature and benefits of their services are interspersed throughout.

All studies are similar in regard to format and methodology. The financial transparency study, for example, conducted interviews with representatives from six NGOs, one state institution and two organisations that manage reusing search and analytical systems <sup>(74)</sup>, while the forestry study interviewed representatives from seven NGOs, one data reuser and one state-sponsored research institute <sup>(75)</sup>. Each report lays out a number of potential benefits of open data in the specified sector and then describes the ways in which reusers in the public and private sectors are currently using data

<sup>(73)</sup> Ministry of Digital Transformation of Ukraine, 2021c.

<sup>(74)</sup> Ministry of Digital Transformation of Ukraine, 2021a.

<sup>(75)</sup> Ministry of Digital Transformation of Ukraine, 2021b.

to reach these goals. The reports conclude with recommendations for continued improvement in the application of open data.

The portal's **analytics dashboard** tracks the creation of new datasets, provides rankings of the keyword tags most frequently attached to datasets and lists the most popular datasets. Since these graphics will not load in archived versions of the site (only headings may be viewed), it is not possible to evaluate the frequency with which these rankings were updated, nor to determine whether popularity was based on visits or downloads. Furthermore, API calls are monitored on a daily basis, which means that the months-long downtime of the portal is clearly visible (see Figure 14). Currently, this statistic is only available at the level of total portal traffic and does not allow for analysis at the level of user groups or data categories.

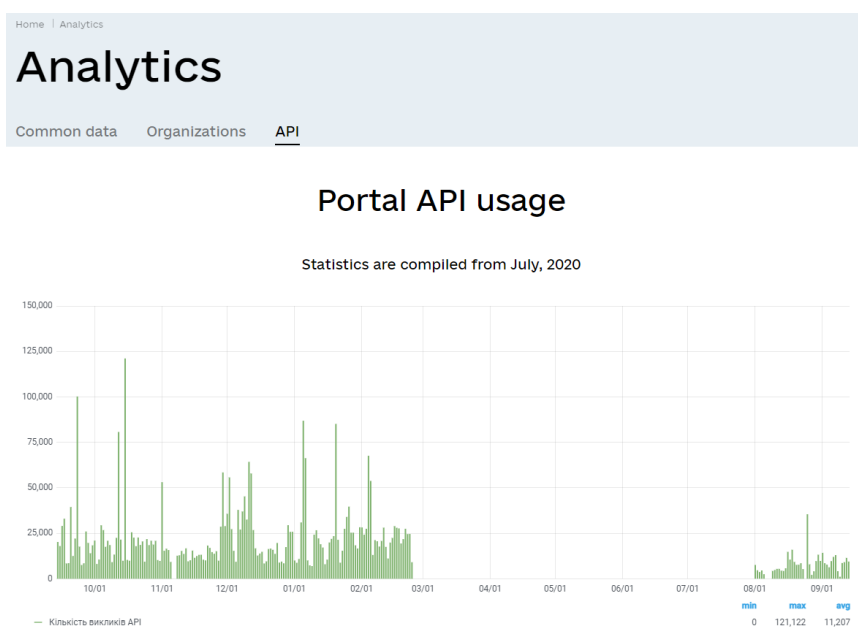


Figure 14. API usage on the national open data portal of Ukraine

## Key findings across national open data portals

Analysis of open data portals at the national level shows that national administrations employ a wide set of strategies to monitor the impact of open data. The presence of a use case repository is the most commonly employed mechanism, although the content and scope of these repositories vary widely.

While use cases mostly give insight into the **path** through which open data impact can be achieved, sometimes information about the more long-term impact is presented. Strong use case sections have the following characteristics.

- They display a wide variety of reuse applications across types, platforms and subject areas.
- They solicit new submissions through forms that encourage detailed description and consistent tagging and formatting.
- They categorise reuse cases appropriately and consistently.
- They allow users to filter cases in accordance with these categorisations.
- Where keyword-based systems are implemented to allow for more specific filtering, they reduce discrepancies and redundancies to a minimum.
- Intended benefits are specified, thereby giving insight into the expected benefits;



- Exemplary or highly impactful reuse cases are featured in dedicated articles that clearly explain the reason(s) for highlighting particular cases.

All but the German, Lithuanian and Polish portals publish **use and user statistics on a dedicated dashboard**, although the number of tracked metrics and the type of metric tracked vary. Strong statistical dashboards provide:

- a curated range of up-to-date statistics;
- a clear specification of the temporal range and update frequency;
- an option for visitors to select a date range from within which to view statistics;
- metrics on dataset popularity by number of individual downloads;
- metrics on the thematic distribution of downloads;
- metrics on the thematic distribution of reuse cases.

Five of the assessed countries (Estonia, Cyprus, Lithuania, the Netherlands and Ukraine) conduct dedicated **studies to assess the impact** of open data. The studies by Cyprus and the Netherlands were also considered in the overview of national studies in Section 1, where the focus is on the applied definitions and typologies of open data impact. The present section investigates the studies more thoroughly.

The methodologies of the five countries vary widely. The Estonian, Cypriot and Lithuanian researchers conducted surveys, while the Dutch and Ukrainian portal studies were based on interviews. The Dutch approach was somewhat more holistic. Its aim was not to produce a report measuring impact within a defined period, but rather to enable continuous monitoring. The conclusions of studies that aim to assess the impact of open data across multiple sectors (or across the broader data landscape) tend to be somewhat limited by the expertise of their respondents. A survey distributed to data reusers, for example, could compile their opinions on the accessibility and value of data, while a survey distributed to public sector agencies could only request that these agencies estimate user satisfaction with the data they had published. Both the Estonian and Cypriot surveys rely to some extent on this kind of proxy assessment of impact <sup>(76)</sup>. Moreover, these surveys generally did not request that respondents provide **evidence** to support their statements. Their conclusions may thus provide more insight into **perception** than into realised or quantified impact.

Soliciting response seems to have posed a challenge to nearly all portal teams. The Estonian study in particular faced very low response rates among public sector agencies, which its authors acknowledged as a limitation in a blog post summarising its conclusions. Conversely, in Ukraine, it seems to have been an intentional choice to interview only a small number of key figures in each policy area.

## 2.2 Local open data portals

The data portals of the cities in Figure 15 will be discussed in this subsection.

---

<sup>(76)</sup> It is not clear whether this is true for the Lithuanian survey, as the methodology and results had not been published at the time of writing this report.

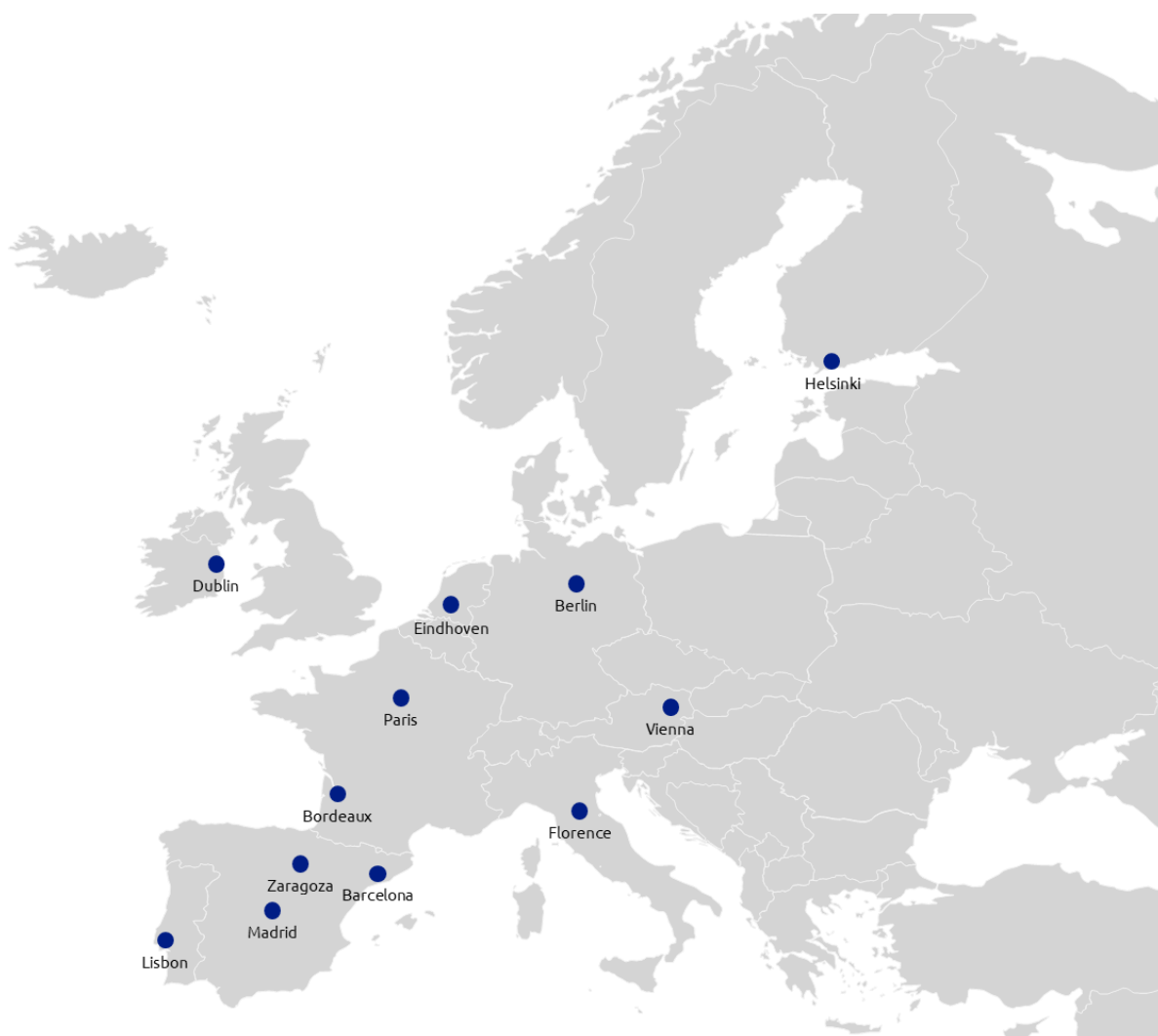


Figure 15. Overview of cities whose open data portal is included in this report

## Barcelona

The [open data portal for the city of Barcelona](#) allows impact assessment to be carried out through a variety of means: a reuse case repository, portal statistics, case stories and interviews with reusers.

The **reuse case repository** takes the form of [visualisations](#) and [applications](#). Six applications and 13 visualisations are currently listed. Each case is briefly described and links are usually provided to consult source datasets (see Figure 16). Visualisations may be consulted by topic: air quality, trees, biking stations and traffic accidents managed by the local police. These visualisations are not consistently cross-listed to the main page, although the descriptions of those that are contain links to the ‘Visualisations by topic’ section. Regardless, this may render it difficult for visitors to locate reuse cases. A form is available for the submission of new visualisations or applications. The purpose of the reuse case needs to be specified using the options ‘Academic’, ‘Social’, ‘Business’ or ‘Other’. However, once added to the repository, cases are not necessarily categorised accordingly. This submission form also specifies that applications will be published for at least 1 month as long as they are deemed to

'work correctly'. In the event that 'any anomaly' is detected, the offending application will be removed with no guarantee of republication after it is fixed.

The screenshot shows the 'Open Data BCN' website, which is the Ajuntament de Barcelona's open data service. The navigation menu includes: ABOUT THIS SITE, DATASET CATALOGUE, PRESENT, DATA VISUALIZATIONS AND APPLICATIONS, STATISTICS, DEVELOPERS, and DIFFUSION PROJECTS. The main content area is titled 'Data visualizations' and includes a sub-header: 'In this section we want to give visibility to the visualizations made with data published in Open Data BCN. You can send us your proposal by filling in the following form.'

Two featured visualizations are shown:

- Most popular and new tendencies for names in the city of Barcelona during 2018**: An infographic titled 'NOMS MÉS POPULARS I TENDÈNCIES 2018 A LA CIUTAT DE BARCELONA' by Raquel Espiña. It features bar charts showing the most popular and new tendencies for male and female names in Barcelona during 2018, based on the dataset '25 Most common baby names in Barcelona, disaggregated by sex'.
- Visualizations of the dataset "Street trees of the city of Barcelona"**: A map titled 'Map of the Street Trees in Barcelona' showing the distribution of street trees in the city.

Figure 16. Data visualisations on the open data portal of the city of Barcelona

Specific uses of data falling under three themes (the environment of the public administration, the academic environment and the professional environment) are highlighted in a 'Case Stories' section. This section also compiles articles pertaining to open data and research using data from the portal. Brief descriptions of these cases are provided, as are links to source datasets.

The portal also features **interviews with data** reusers. Between February 2015 and April 2016, nine interviews were conducted with both individuals and representatives of reusing companies. The length and format of these interviews vary widely, but reusers are generally asked to explain the nature and purpose of their data usage and to provide their thoughts on the value of open data, the process of reuse and the quality of the portal.

The portal has a comprehensive use and user **statistics dashboard**. Metrics tracked include site language, user country of origin, user device type and the distribution of datasets across the following themes: administration, city/services, population, territory and economy/enterprise. Rankings of the five most-visited datasets, most-downloaded datasets by IP address and most-downloaded datasets by absolute value from the past 15 days and cumulatively since February 2017 are also provided. Visitors to the dashboard may view the number of currently open requests for data in graph format.

## Berlin

The impact of the [open data portal for the city of Berlin](#) can be assessed through reuse cases and a dataset containing use statistics.

The portal lists 66 **reuse applications**, some of which are ‘older than the portal itself, but are listed to show what developers and other interested parties can produce using freely accessible data’ (77). Case subpages contain brief descriptions, links to source datasets and ‘tags’ sections, where tags may refer either to subject matter (e.g. ‘Christmas’) or to type (e.g. ‘App’). This increases the likelihood of relevant applications appearing in searches, but since there is no option to filter or search the case section exclusively, visitors must perform indiscriminate keyword searches of the website. Pages under the ‘Interaction’ section and datasets are also tagged, so a query for ‘bicycle’ will return applications, datasets and articles pertaining to cycling. Reuse cases are also not categorised and cannot be filtered directly by tag or type on the ‘Applications’ page.

Visitors may submit new cases for inclusion through a general contact form, the *Anwendung hinzufügen* option visible on the left in Figure 17.

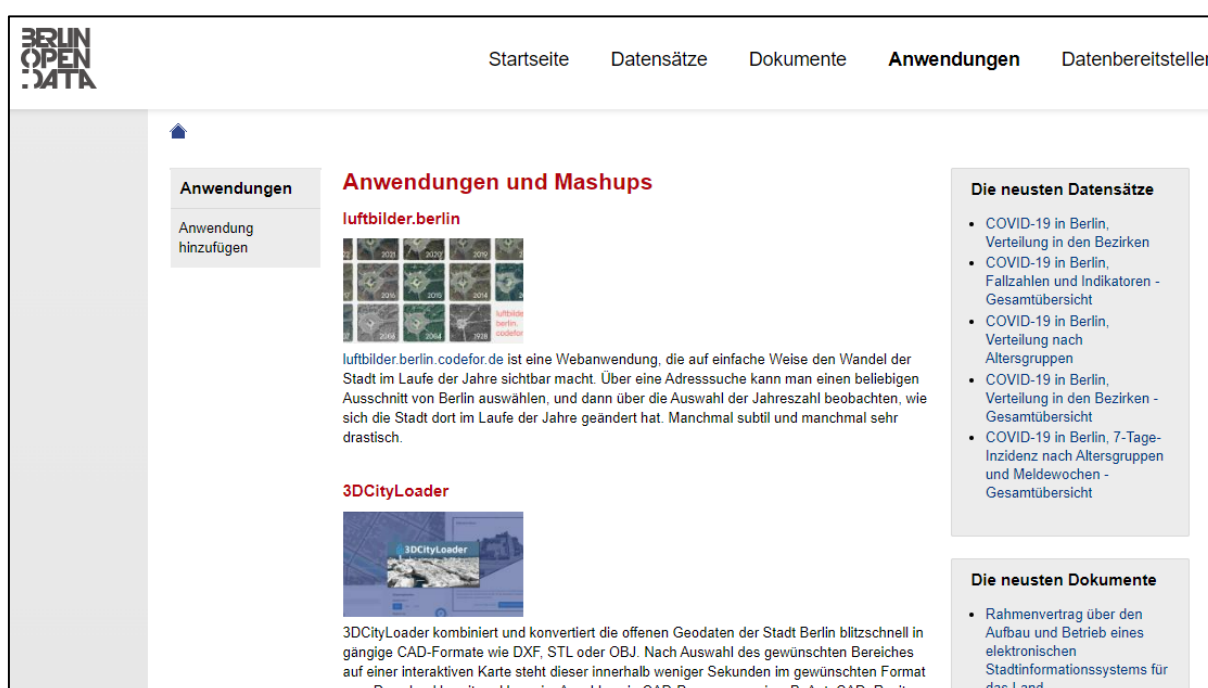


Figure 17. Reuse case section on the open data portal of the city of Berlin

The portal highlights certain datasets that have produced **multiple reuse applications**, under the ‘Interaction’ section (78). It also **highlights datasets that are particularly relevant to current events**, such as a profile at the beginning of the COVID-19 pandemic on a dataset compiling businesses that offer delivery services. The portal publishes **use statistics** in CSV format, which are listed in the dataset repository and updated monthly. Although the portal tracks impressions and visits for all data records and for the general domain, it does not track downloads, so little may be inferred from reuse potential.

Berlin is currently in the process of developing a **new strategy for opening data** (79). The first phase of development includes a 24-question survey available to the public. The survey’s questions primarily concern respondents’ perception of the importance of open data and, for those who have reused data from the portal, the ease of the process. The full text of selected questions that are relevant to measure

(77) Translation by authors. Original text listed on the homepage: ‘Einige dieser Beispiele sind älter als dieses Portal, wir haben Sie hier trotzdem aufgeführt um zu zeigen, was Entwickler und Interessierte aus frei zugänglichen Datensätzen erschaffen können’.

(78) <https://daten.berlin.de/interaction> (in German)

(79) <https://strategie.odis-berlin.de/> (in German)

open data impact may be viewed in Annex IV– Impact measurement on national and local open data portals. Participants were not required to answer all 24 questions. While most received approximately 50 responses, fewer participants chose to respond to the survey’s three open-ended questions, which received 31, 32 and 30 responses respectively. As of March 2022, the survey phase of the process was concluded, after which the city organised stakeholder workshops on the themes of administration, civil society, business and science. The results of the workshops will be integrated into the recommendations for the new Berlin open data strategy (which haven’t been published yet).

## Bordeaux

The [open data portal for the city of Bordeaux](#) includes a **reuse case repository** with a brief description of each case (see Figure 18). Cases are tagged and may be filtered by theme (e.g. citizenship), territory (e.g. Bordeaux Metropolis) and type (either application or visualisation). The option to filter by theme may be seen at the top left. Cases are also tagged with the theme of their source dataset (e.g. displacement), although no option is provided to filter cases accordingly. In Figure 18, ‘Les événements talençais’ is categorised under ‘société’ and ‘culture, sports et loisirs’, but there is no fourth menu allowing a visitor to view all reuse cases with the latter tag. Dataset themes are not attached to specific case themes. Links to source datasets are also not provided, even though the form for submitting new cases requests that submitters specify the data that they used from the portal. The portal’s forum section includes a board for discussing reuse cases. Although this could provide a space for user engagement and feedback, since December 2011, only five posts have been made, all by developers promoting their applications.

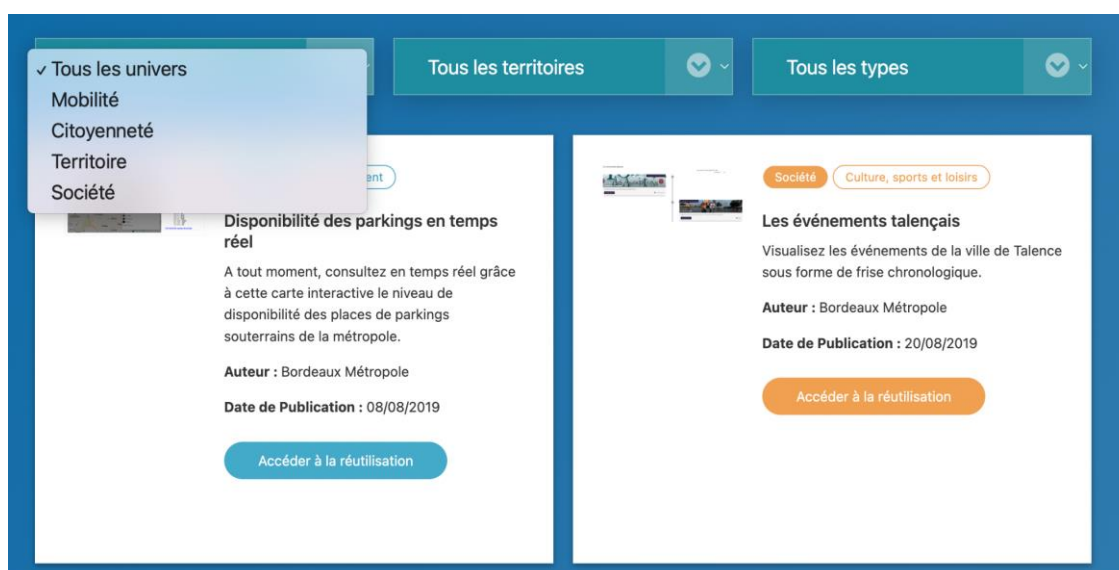


Figure 18. Reuse case section on the open data portal of the city of Bordeaux

## Dublin

The city of Dublin assesses the impact of its [open data portal](#) through a **statistics dashboard**<sup>(80)</sup>. The statistics compiled relate primarily to the quality and source of the datasets available on the portal. Tracked metrics include datasets added each month, datasets added each month by each local authority and dataset compliance with formatting and metadata requirements. The total number of users, average number of pageviews per session and the share of API-accessed datasets are equally

<sup>(80)</sup> <https://data.smartdublin.ie/stats>

provided. However, no detailed metrics at the level of data categories or types of users are available. The most relevant statistics from an impact perspective are the rankings of the 10 most-searched keywords and most-viewed datasets, since they provide some insight into the interests of potential reusers. Although some of the listed URLs appear to be duplicates, when clicked on, they lead to separate dataset subpages. Since the number of downloads is not tracked, it is difficult to ascertain the reuse potential. The temporal range of these statistics is unclear.

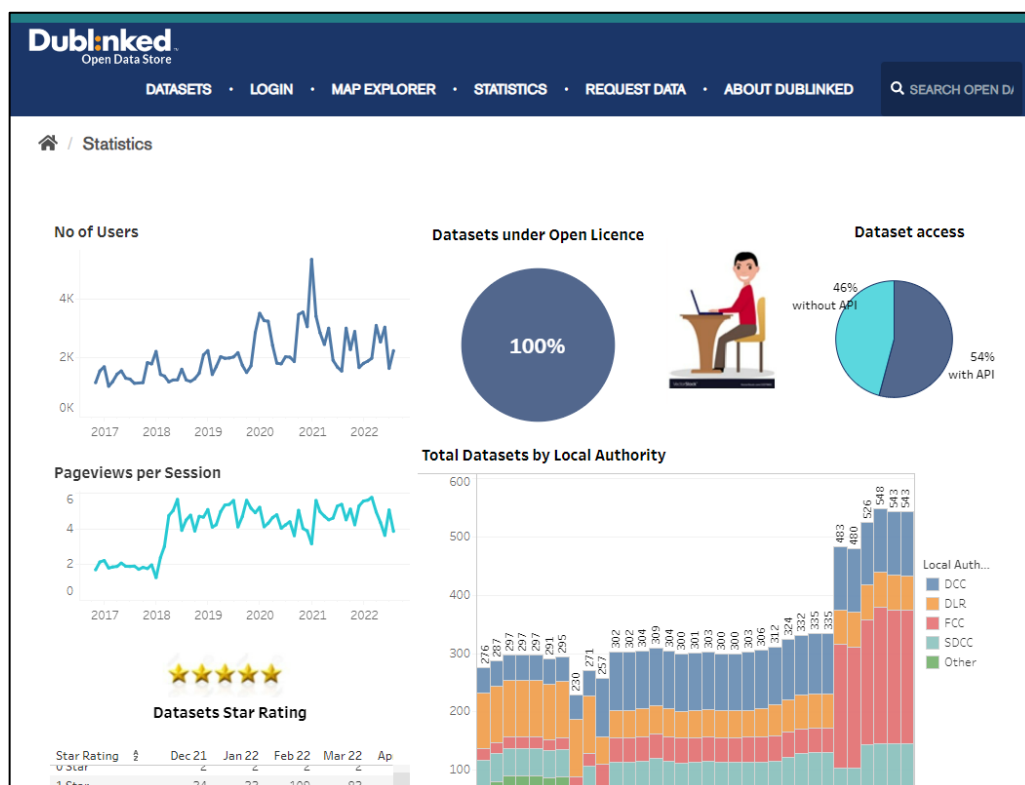


Figure 19. Statistics dashboard on the open data portal of the city of Dublin

## Eindhoven

The primary means by which the impact of the [open data portal for the city of Eindhoven](https://onderzoek.eindhoven.nl/onderzoek-en-informatieproducten) may be assessed is through its ‘research and information products’ page, which links to several **cases of reuse** <sup>(81)</sup>.

A brief description is provided for each reuse case. Cases are categorised and may be filtered by theme (e.g. ‘economy, work & income’). Visitors may also search the page by keyword. These options are visible on the left side of Figure 20. Once a visitor has selected a theme, a secondary menu listing subthemes will be displayed, allowing further refining. The 25 examples of reuse provided are all dashboards, factsheets, maps and reports by the municipal administration itself. There are no options for external reusers to submit reuse cases for inclusion.

<sup>(81)</sup> <https://onderzoek.eindhoven.nl/onderzoek-en-informatieproducten> (in Dutch with integrated machine translation available)

The screenshot shows the Eindhoven open data portal interface. At the top left is the Eindhoven logo and the text 'EINDHOVEN'. At the top right is a 'Translate' button. Below the header, there is a breadcrumb trail: '/ Research and information products'. A paragraph states: 'On this page you will find the results of studies and information products per theme. This can be in the form of dashboards, fact sheets, maps and reports.'

On the left side, there are three filter sections:

- Keywords:** A search input field.
- Theme:** A list of checkboxes:
  - Population & governance
  - Economy, work & income
  - Physical living environment
  - Social living environment
- Topic:** A list of checkboxes:
  - Population
  - Diversity
  - Youth
  - Environment & sustainability

Two featured data cards are displayed on the right:

- Place waste separately:** Includes a map with red location markers and the text: 'You don't want to walk unnecessarily far to separate your waste. You are therefore looking for the locations of underground and above-ground waste separate places in your area! These locations are...'. A right-pointing arrow is at the bottom right.
- Labor force:** Includes a photo of people working at a table and the text: 'The working population consists of all Eindhoven residents between the ages of 15 and 75 with a job. Or, if not, have recently looked for paid work and are available for it.'. A right-pointing arrow is at the bottom right.

Figure 20. Reuse case section on the open data portal of the city of Eindhoven (machine translated)

While the portal has no statistics dashboard, the homepage provides a ranking of the top five most-downloaded datasets. This metric appears to be cumulative, but no temporal range is specified.

## Florence

The [open data portal for the city of Florence](https://opendata.comune.fi.it/statistiche) provides a **statistics dashboard** as a means of assessing impact <sup>(82)</sup>.

Three metrics are tracked: the country of origin of site visitors (see Figure 21), which may also be seen in map format, the top 10 most-downloaded datasets and the device used to access the portal (e.g. tablet). Unlike the Dublin portal, the Florence portal does not provide links to datasets in its ranking. Visitors must search independently for these datasets if they would like to view or download them. Statistics may be viewed cumulatively or for any month since December 2018, although selecting the cumulative option for the country-of-origin metric distorts the graphic.

<sup>(82)</sup> <https://opendata.comune.fi.it/statistiche> (in Italian)

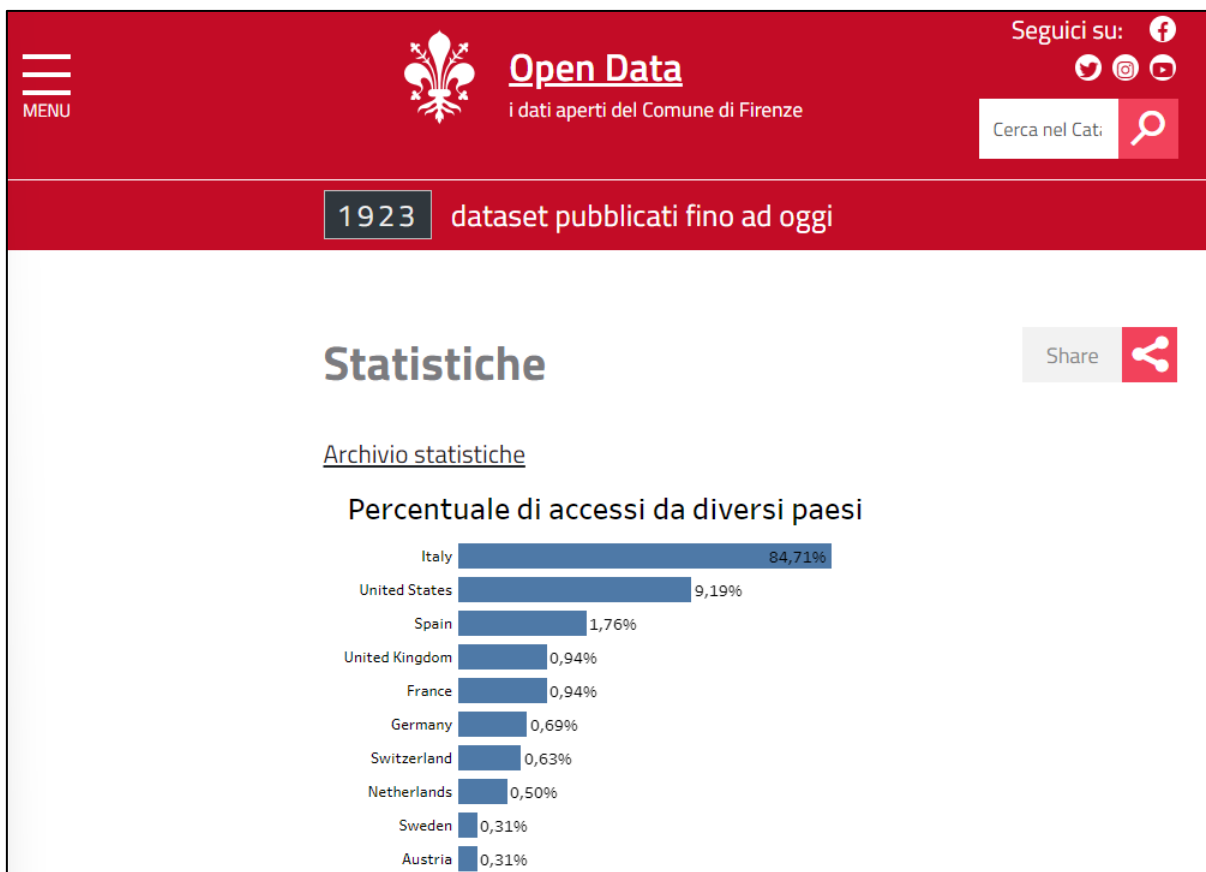


Figure 21. Ranking of visitors' country of origin on the open data portal of the city of Florence

## Helsinki region

The city of Helsinki does not have a municipal open data portal but has joined forces with three other Finnish municipalities – Espoo, Vantaa and Kauniainen – to create an [open data portal for the Helsinki region](#). The portal compiles reuse cases and tracks portal use statistics, both of which allow for impact assessment.

The portal lists 306 reuse applications in a **showcases section** which may be accessed from the homepage via either a menu or an updating widget displaying the total number of listed cases <sup>(83)</sup>. The subpage for each case contains a brief description and links to source datasets. Cases are tagged and may be filtered by type (e.g. visualisation) or platform (e.g. Android). The menu of type options is visible on the left in Figure 22. Some redundancies are present in the platform tagging system: 'iOS' and 'iPhone' are tagged separately, and a second iOS tag appears to have been erroneously attached to two web-based applications. The portal also employs a keyword-based filtering system where keywords may refer either to the subject (e.g. 'public transport') or the nature of the application (e.g. 'journey planner'). There is no option to see all keywords or to filter by keyword on the main reuse case page, but navigating to a case subpage and selecting a tagged keyword will lead a user to a list of all applications labelled with that phrase. That list may subsequently be filtered by theme or platform. Keywords seem to have been applied with relative consistency, although the number of tags varies widely between applications. Both **new applications and new ideas for applications may be submitted**

<sup>(83)</sup> [https://hri.fi/data/en\\_GB/showcase](https://hri.fi/data/en_GB/showcase)



through provided forms. Six ideas for open data-based applications are listed on a second page, but no data is provided on whether these suggestions have led to any actual development.

Figure 22. Reuse case section on the open data portal of the Helsinki region

The portal's dynamic **site analytics dashboard** allows the user to visualise selected statistics over a custom time period. The dashboard provides: 1) a ranking of the top 15 datasets by views, 2) a timeline of the total number of views, 3) a thematic distribution of available datasets, 4) a distribution of available file formats, 5) a distribution of datasets according to publisher, 6) a timeline of total available showcases and 7) a type distribution of reuse cases (e.g. visualisations, data journalism and city tools).

## Lisbon

The national [open data portal for the city of Lisbon](#) allows for impact assessment by showcasing applications and publishing a selection of data-driven visualisations and dashboards created by the city itself.

Nine **reuse examples** are highlighted on the portal, with a brief description of each case <sup>(84)</sup>. Eight of the nine cases were presented at Smart Open Lisboa 2016, a conference for start-ups and corporate partners, where one-minute video interviews were conducted with each application developer; these interviews are included in the section. Two video thumbnails are visible in Figure 23. There does not appear to be an option for visitors of the portal to submit new reuse cases for inclusion.

<sup>(84)</sup> <https://lisboaaberta.cm-lisboa.pt/index.php/pt/apps-e-analitica/apps> (in Portuguese)

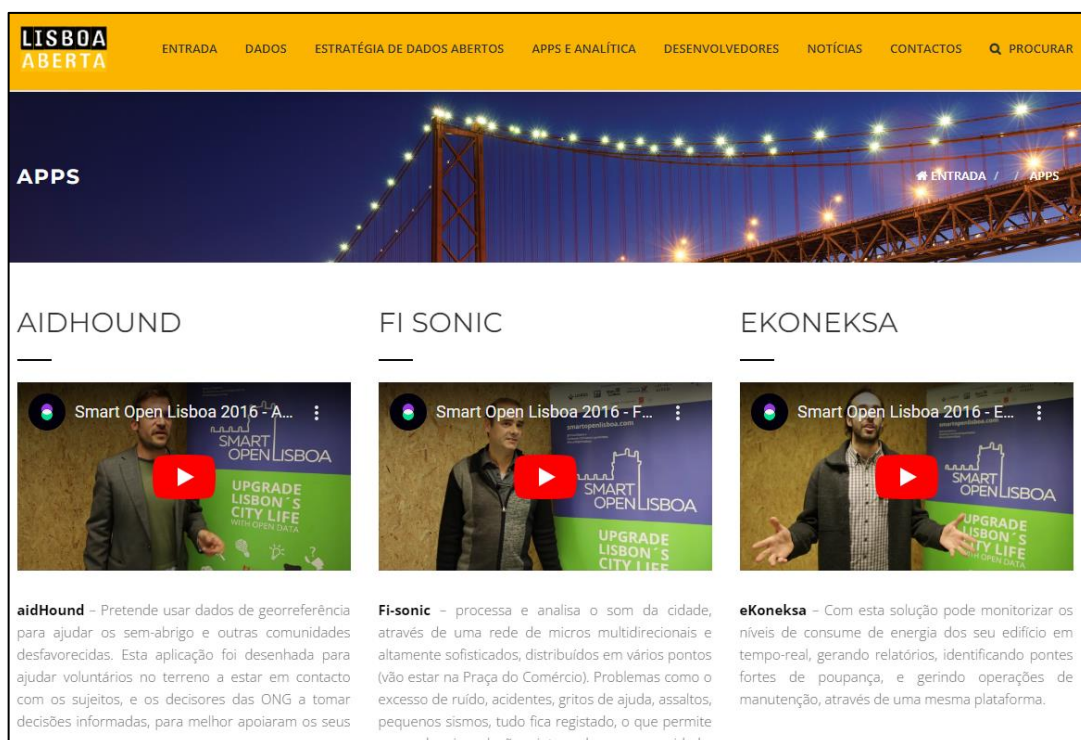


Figure 23. Reuse case section on the open data portal of the city of Lisbon

The portal provides links to six **thematic dashboards**, which are not always up to date, featuring open data visualisations created by the city of Lisbon <sup>(85)</sup>. The themes are: environment, events on public roads, tourism, population, public works and education.

## Madrid

The impact of the [open data portal for the city of Madrid](#) may be assessed by means of its reuse case section and to a lesser extent through a statistical dashboard.

The **applications repository** lists 35 examples of reuse with a brief description <sup>(86)</sup>. New cases may be submitted for inclusion through a form which requests that submitters specify the datasets used from the portal. Links to source data are not provided for individual cases, but case descriptions sometimes provide some information regarding the datasets. The description of the application 'Madrid Metro | Bus | Cercanías', for example, visible in Figure 24, specifies that 'open data on buses from EMT and BiciMAD' was used.

<sup>(85)</sup> <https://lisboaaberta.cm-lisboa.pt/index.php/pt/apps-e-analitica/analitica> (in Portuguese)

<sup>(86)</sup> <https://datos.madrid.es/portal/site/egob/menuitem.400a817358ce98c34e937436a8a409a0/?vgnextoid=994612b9ace9f310VgnVCM100000171f5a0aRCRD&vgnnextchannel=994612b9ace9f310VgnVCM100000171f5a0aRCRD&vgnnextfmt=default> (in Spanish)

**Tu ciudad más cerca**  
Gracias a nuestra plataforma de datos abiertos podrás encontrar todos los datos de Madrid que necesitas para tu proyecto

En portada | Acerca de Datos Abiertos | Catálogo de datos | **Colabora** | Visualizaciones

Lo más visto ⓘ ActivadesMadridBot / FuentesMadridBot

Colabora > Aplicaciones

**Aplicaciones**

Propuestas recibidas

Proponer un conjunto de datos

Los conjuntos mejor valorados

**Aplicaciones**

Periodismo de datos

Informar sobre una aplicación realizada

Registro de reutilización

En esta sección se presentan ejemplos de aplicaciones que han desarrollado personas y empresas reutilizando los conjuntos de datos publicados del portal de datos abiertos del Ayuntamiento de Madrid.

Para que publiquemos tus aplicaciones, utiliza el formulario [Informar sobre una aplicación realizada](#) con conjuntos de datos del portal de datos abiertos.

Total: **35** Mostrados: **1-10**

Madrid Metro | Bus | Cercanías

Aplicación que facilita la consulta de todo el transporte público de Madrid: tiempos de espera, rutas, paradas y horarios de autobuses urbanos, interurbanos, Metro, Cercanías y BiciMAD. Esta aplicación utiliza, entre otros, los datos abiertos de autobuses de la EMT y de BiciMAD.

Figure 24. Reuse case section on the open data portal of the city of Madrid

A **statistical dashboard** features three graphs showing 1) the number of users, sessions and page visits per year since 2017; 2) the evolution in the monthly average of the same metrics; and 3) the monthly evolution in number of downloads for the years 2020, 2021 and 2022 <sup>(87)</sup>. Similar to other open data analytics, no detailed metrics related to specific themes or user types are available.

## Paris

The primary mechanism for assessing the impact of the [open data portal for the city of Paris](#) is its reuse case repository <sup>(88)</sup>. It is accessible through a widget on the homepage that displays the three most recently added cases. A brief description is provided for each case. Both applications and academic studies of data are included in this section. New reuse cases can be suggested for inclusion through the 'information' tab of datasets, ensuring that all listed cases provide links to source data. The subpages for datasets link to all related reuse cases, enabling some assessment of reuse potential.

<sup>(87)</sup>

<https://datos.madrid.es/portal/site/egob/menuitem.400a817358ce98c34e937436a8a409a0/?vgnnextoid=d11ce2e5b6801610VgnVCM1000001d4a900aRCRD&vgnnextchannel=d11ce2e5b6801610VgnVCM1000001d4a900aRCRD&vgnnextfmt=default> (in Spanish)

<sup>(88)</sup> <https://opendata.paris.fr/pages/les-reutilisations/> (in French)

Figure 25. Reuse case section on the open data portal of the city of Paris

Although the portal has no statistics dashboard, a ranking of the top five datasets by number of downloads may be viewed on the homepage.

## Vienna

The [open data access point for the city of Vienna](#) provides visitors insight into the reuse of data from Vienna through the reuse case repository of the Austrian national open data portal <sup>(89)</sup>. In one click, all 324 Vienna-based reuse cases are listed. A brief description is provided for each case. Cases are categorised and may be filtered by type (e.g. 'visualisation'), platform (e.g. 'Symbian'), data origin (e.g. 'Federal Geological Survey') and theme (e.g. 'education and research'). There are some discrepancies in the platform tagging system. As is visible in Figure 26, 'Windows' and 'Windows 8' are tagged separately. The filters menu shows the number of datasets falling under each category. In the absence of a statistics dashboard, this allows the reuse potential of the data provided by municipal organisations to be estimated to a certain extent. Reuse case subpages consistently link to the datasets used. The form through which visitors may submit new applications for inclusion mandates the provision of source data. Submitters must specify the type and platform from drop-down menus. It appears that themes may be automatically assigned based on the theme of the first source dataset.

<sup>(89)</sup> [https://www.data.gv.at/applicationdatapublisher/stadt-wien/?post\\_type=anwendungen&applicationsystem=0&applicationtype=0](https://www.data.gv.at/applicationdatapublisher/stadt-wien/?post_type=anwendungen&applicationsystem=0&applicationtype=0) (in German)

The screenshot shows the 'Anwendungen' (Applications) section of the data.gov.at portal. At the top, it says 'Nationalrat einstimmig für erleichterte Weiterverwendung von Daten im öffentlichen Sektor'. Below the navigation bar, there's a heading 'Anwendungen' and a sub-heading 'Sie haben auch eine Anwendung basierend auf offenen Daten erstellt und möchten diese hier vorstellen? Dann informieren Sie uns bitte darüber mit Hilfe des Formulars Anwendung einreichen.' (Seite 1 von 33).

On the left, there are filter sections:
 

- Filter**: Anwenden, Zurücksetzen
- 324 Einträge gefunden**
- Typ**:
  - Datenaufbereitungstool (105)
  - Konzept (37)
  - Mobile App (237)
  - Partizipation (19)
  - Tool (10)
  - Tutorial (12)
  - Visualisierung (275)
  - Web App (242)
  - Wissenschaftliche Publikation (21)
- Betriebssystem**:
  - Android (153)
  - BlackBerry (36)
  - Browser www (418)
  - Firefox OS (2)
  - iOS (139)
  - Linux (23)
  - MacOSX (1)
  - OSX (14)
  - Sonstiges (11)
  - Symbian (3)
  - Watch OS (5)
  - WebOS (1)
  - Windows (31)
  - Windows 8 (21)

Three application cards are shown:
 

- Wie Weit?**: Eingereicht von Stefan Wagner. Description: 'Wie weit kommt man innerhalb von 30 Minuten mit den öffentlichen Verkehrsmitteln der Wiener Linien und ÖBB? Basierend auf ÖBB [...]'
- MapColPal – your pal for cartographic color palettes**: Eingereicht von Valerian Lange. Description: 'MapColPal ist eine Web-Anwendung, die beim Erstellen von Farbpaletten für die kartographische Visualisierung hilft. Daten von data.gov.at werden als Beispieldaten [...]'
- Wasser in Wien**: Eingereicht von Mathias Hackl. Description: 'Wasser in Wien – Finde die nächsten öffentlichen Wasserstellen, Schwimmbäder, Badestellen, Toiletten, Trinkbrunnen, Sprühnebel Duschen und Hundetrinkbrunnen in Wien.'

Figure 26. Reuse case section on the open data portal of the city of Vienna

## Zaragoza

The [open data portal for the city of Zaragoza](#) provides two mechanisms for assessing impact: a reuse case section and a statistics dashboard.

The portal lists 55 reuse cases in its **reuse case repository** <sup>(90)</sup>. Although the true number of reuse cases is displayed on the main reuse page, the homepage provides an estimate in a widget. This widget also displays the total number of datasets published through the portal, an estimate of daily queries to dataset pages and an estimate of the number of registered reusers. A brief description of each case is provided on the main page. Links to case subpages currently appear to be currently non-operational. Cases are categorised and may be filtered by platform (e.g. Android) and device (e.g. mobile). The portal also uses a more specific keyword-based tagging system. Like other keyword-based systems used in national portals, a number of discrepancies and redundancies are present that may complicate searching. As shown in Figure 27, for example, 'Movilidad' is tagged separately from 'movilidad', and the tags yield different sets of cases when selected. 'Android' and 'mobile' both appear as general tags when the other categorisation options should provide this information. Visitors to the portal must register as reusers before submitting new applications for inclusion.

<sup>(90)</sup> <https://www.zaragoza.es/sede/servicio/aplicacion/> (in Spanish)

The screenshot shows the 'Datos Abiertos' portal for Zaragoza. The header includes the logo, navigation links for 'API REST v2', 'SPARQL', 'SOLR', and 'Vocabularios', and a search bar. The main content area is titled 'Buscador de aplicaciones' and displays search results. A sidebar on the left lists various themes like 'Movilidad', 'Equipamientos', 'Zaragoza', 'bus', 'movilidad', 'transporte', 'zaragoza', 'Actividades', and 'Eventos'. The search results section shows a card for 'ZGZ' (Zaragoza Tránsito) with a QR code and a description: 'Descubre cuánto tiempo de espera le queda al tranvía de Zaragoza. Busque la parada en el mapa y conoce cuánto tiempo queda. Requisitos iOS 14.0'. The author is listed as 'Javier Gofí Puerta'.

Figure 27. Reuse case section on the open data portal of the city of Zaragoza

The portal has a section with **use statistics** accessible from another link on the homepage <sup>(91)</sup>. This page tracks three metrics, all over the course of the past 30 days: the formats in which data files were accessed (e.g. CSV), the HTTP methods through which they were retrieved by the API (e.g. GET) and the total number of dataset consultations per day.

## Key findings across local open data portals

The analysis of local open data portals shows that municipal administrations employ a mix of methods to assess the impact of open data. Among these portals, **compiling examples of reuse** was the most common mechanism for impact assessment: 10 of the 12 assessed portals have some form of case repository, although the number of applications listed and the amount of detail provided for each application vary widely.

Like those found on national portals, many of the reuse case repositories employ categorisation systems, usually based on fixed theme and application typologies or broader keyword-based tagging. Where the latter is employed, the usual discrepancies and redundancies are present. In the absence of active tag moderation, these issues seem to be inherent to keyword-based categorisation. On the Helsinki Region portal, where keyword tags are assigned by portal staff and not submitted independently by reusers, these tags are applied with much more consistency across cases. While descriptions of reuse applications are consistently provided, links to source data were comparatively rarer at the local level.

<sup>(91)</sup> <https://www.zaragoza.es/sede/servicio/datos-uso/> (in Spanish)

Local portals tend to provide **narrower ranges of use and user** statistics than their national-level counterparts. Dataset popularity rankings by downloads and visits are commonly included. Several portals also track the distribution of reuse cases by thematic category.

**Impact studies are comparatively scarce** at the local level. Only Berlin conducted a survey-based assessment akin to that undertaken by some national portals. The German capital's research is somewhat limited in its findings though, since it mainly focuses on data access and intended reuse rather than quantified or proven impact. Its questions may provide some insight into portal quality and the distribution of reusers across sectors, although the latter could be assessed more comprehensively through the analysis of the portal's reuse case repository.

## 2.3 Conclusion: considerations for the development of an open data impact assessment methodology

Measuring the impact of open data is not only about knowing the effects on the economy and society as a whole, but also about understanding through what **paths** these are achieved. By providing use(r) statistics, use case stories, user research and analytical reports, open data portals across Europe provide bigger and smaller breadcrumbs to measure the path to impact.

The **main challenge** ahead is to find ways to connect these data points to come to genuine insights about the impact of open data. How can metrics from download statistics, compelling reuse cases and reuser surveys help understand the effect of open data in the social, political, economic and environmental domains?

Based on the observed strengths and weaknesses of the analysed portals, the following issues should be taken into consideration for the development of an impact assessment methodology for European open data.

- **Connecting indicators on short-, medium- and long-term effects to measure the path to impact**

The **short-term effects** of open data publication have to do with consultation of and access to the data. Downloads and views are popular metrics regarding such short-term impact. For impact measurement at the meso or macro level, it is crucial to consider relevant categorisations for such indicators. Both technical aspects of data access, such as consulted data formats, used API methods and thematic areas (e.g. mobility and health) can be used, but will only help to get insight on the path to impact if they can be connected to information on the medium- and long-term effects of open data. The production of vanity statistics, such as the number of views of a particular dataset not tied to reuse applications, should be avoided as they give a false impression of impact. The analysis shows that only some use cases clearly state the dataset(s) used. In order to measure the impact of a certain dataset, the link between the dataset and the corresponding use case(s) should be clear. This can be achieved by adding a standard field for the source dataset to the use case submission form, for example.

The **medium-term effects** of open data become visible when they are reused. Reuse case descriptions can provide valuable qualitative insights at the micro level, especially when the

datasets used and the experienced benefits are equally documented as part of a value chain approach. To obtain more quantitative insights at the meso level, it is essential to standardise use case descriptions and leverage the automated generation of cross-case indicators where possible. Such analysis may for instance reveal which thematic categories are dominant in the development of consumer apps. Smart categorisations and filtering options will further serve such analyses. Inconsistent tagging by reusers is a prevalent challenge, as is the submission of reuse cases in the first place.

The **long-term effects** of open data relate to the production and uptake of data-driven products and services. Such impact remains difficult to measure and is predominantly considered in terms of estimated impact. To be able to do so, it is necessary to consistently document realised benefits and beneficiaries in reuse cases, and carry out recurring impact studies.

- **Distinguishing between estimated and realised impact**

On several portals, a mix of indicators relating to the realised and estimated impact of open data was found. Statistics on data requests, for instance, may provide insight into the demand for a particular dataset, its data category and possibly its reuser category<sup>(92)</sup>. Such insights, complemented by economic projections and expert opinion research, could help to make estimations about future trends in open data use and impact. Regarding realised impact, portals also include statistical and qualitative information tracking the effects of specific datasets or categories of datasets since publication, information tracking the movement of open data from publication to reuse and finally uptake of data-driven services in an impact domain.

- **Measuring impact at the micro, meso and macro level**

The methods in use to measure open data impact suggest that this can be done at multiple levels. At the micro level, it may take the form of tracking the impact of a particular dataset through its specific reuse(s) and subsequent uptake or tracking the impact of a specific reuser (see also the next section on open data intermediaries). At the meso level, open data impact can be understood as the effect of a technical, thematic or other category of datasets on a group of data reusers and end users. At the macro level, it entails understanding the long-term and wide-scale effects of open data practices on the economy and society as a whole. It should be noted though that more datasets, more re-users and more end users does not automatically mean a stronger impact. An influential data reuser, such as a data intermediary or an influential end user such as a policymaker, may be able to generate impact at the macro level.

- **Ensuring an adequate frequency of measurement**

Short-term effects may be measured on a more continuous basis, leveraging automatically generated statistics (see also Section 4). Updating use cases featured on open data portals is important to monitor medium-term effects and impacts at a micro scale. This is especially important, as the benefits of the developed data visualisations, mapping tools and other types of reuse only become apparent after some time has passed. Studies dedicated to assessing the more

---

<sup>(92)</sup> Publications Office of the European Union and European Data Portal, 2022.



long-term and wide-scale impacts of open data benefit from being held on a regular basis rather than a one-off exercise.

- **Standardisation of indicators**

A key challenge is the lack of standardisation. This is true for use(r) statistics, the description and categorisation of reuse cases, and dedicated impact studies. Standardising key metrics and tagging systems across European data portals would greatly improve comparability. Such standardisation efforts may include suggesting keywords to reusers when submitting a showcase for inclusion on the portal and considering comparable metrics such as the ratio between use cases and number of available datasets.

### 3. Measuring the impact of open data intermediaries

The path to social, political, economic or environmental impact does not always go directly from the data provider to the end user. For instance, the impact of COVID-19 data (e.g. infections, death tolls, testing capacity) on citizens' perceptions and evaluations of government health policies more likely happened through aggregations, analysis and visualisations produced by intermediaries than through direct downloads of the source data and interpretation by citizens themselves.

Open data intermediaries form an often-necessary **bridge** between data providers and end users. The latter group may lack the awareness, skills or trust in the original data source to directly access and reuse the data that is relevant to their needs. With the increased availability of open data and thereby complexity of the open data landscape, the role of open data intermediaries has become ever more significant to ensure that the publication of open data leads to the generation of social, political, economic and environmental impact by end users.

Based on a review of academic and policy research, this section of the report will discuss the role of open data intermediaries and how they process data from original sources to make it accessible to other actors for reuse (downstream data), and will propose indicators to measure their impact. The source material used for the analysis includes key open data intermediary case study reports <sup>(93)</sup>, conceptual analyses <sup>(94)</sup> and quantitative studies on open data intermediaries <sup>(95)</sup>.

Section 3.1 will explain what open data intermediaries are and why they matter when measuring open data impact. Then, Section 3.2 will map open data intermediaries in the data value chain, and Section 3.3 will provide an overview of what research says about indicators for downstream open data reuse. Lastly, Section 3.4 will conclude with the considerations of open data intermediaries for the development of an open data impact assessment methodology.

#### 3.1 What are open data intermediaries and why do they matter?

##### Defining open data intermediaries

Open data intermediaries, sometimes called infomediaries, are stakeholders who are '(i) positioned at some point in a data supply chain that incorporates an open dataset, (ii) positioned between two agents in the supply chain, and (iii) [who] facilitate the use of open data that may otherwise not have been the case' <sup>(96)</sup>. In many cases, the intermediary action of such stakeholders goes further than merely pointing an end user to a dataset that is relevant to their purposes. They tend to make open government data more accessible and useful for end users, through added analysis, combination with other data sources, and visualisation or publication <sup>(97)</sup>.

Between what do they intermediate, exactly? The literature is rather ambiguous when it comes to this question. On the supply side, for instance, an open data portal can be considered an intermediary between the data holder (e.g. the specific public agency or department) and the user. On the demand

---

<sup>(93)</sup> E.g. Van Schalkwyk et al., 2015.

<sup>(94)</sup> E.g. Janssen and Zuiderwijk, 2014.

<sup>(95)</sup> E.g. ASEDIE, 2021.

<sup>(96)</sup> Van Schalkwyk, et al., 2015.

<sup>(97)</sup> Janssen and Zuiderwijk, 2014.

side, the intermediary can be a commercial data repository such as Google Public Data explorer or a website providing visualisation of data such as Worldometer for anyone to browse.

For the scope of this study, we exclude from the definition of open data intermediaries:

- official public open data portals such as data.europa.eu;
- websites primarily providing visualisation, storytelling or other data products, such as Worldometer, rather than data for reuse;
- infomediaries, who typically focus on data products and information services based on open data, as the suffix 'info-' suggests.

In other words, the primary demand side of an open data intermediary is the **data reuser**, not the data service end user. The primary output is **data**, not data products. Admittedly, this requires careful judgement, as there is no clear division between the two.

In addition, the scope only covers open public data, not personal, business-held or any non-open data. It considers intermediation regarding data resources generated or commissioned by public institutions in the sense of the open data directive and excludes personal and non-personal data resources provided by businesses, citizens or other societal actors in the sense of the Data Governance Act <sup>(98)</sup>.

Table 3: Example of placement of open data intermediaries in the open data value chain

Data holder	Data portal	<b>Open data intermediaries</b>	Data reuser	Data service end user
Any public body	Government open data portal, e.g. data.europa.eu	<b>e.g. Google public data explorer</b>	e.g. Worldometer	Any citizen, business, societal actor or public body

### Why are intermediaries important

The open data movement has evolved from a focus on publishing as much possible to a more nuanced and fine-tuned approach aimed at increasing open data impact. This is visible for instance in the selective approach of the open data directive in favour of high-value datasets. The reason is simple: open data reuse is harder than originally thought and requires more careful work on quality and usability.

Open data intermediaries play an important role in this regard, as they make open data more accessible, usable and useful to reusers. Typically, they have a greater focus on the needs of their specific users. They aggregate data from different sources and they carry out some form of control over the data quality. It is difficult to quantify the importance of such intermediaries, but there is evidence that such services attract more reusers than the original open data provided by governments. For instance, Van Schalkwyk et al. (2015) find that both non-profit and commercial services are far more used than the original government service, the former however remaining a necessary source of data for the latter.

Better understanding the role of intermediaries can also have important implications for policy. The more important such intermediaries are, the more government open data initiatives should take into

<sup>(98)</sup> European Commission, 2019 and 2022.

account their needs, avoiding competition and pursuing maximum synergy, for instance by investing in API access rather than human user-interfaces.

Based on the cases mentioned in the literature review (Annex V – open data intermediary case study overview), a typology of open data intermediaries was developed, taking into account six dimensions.

Table 4: Typology of open data intermediaries

Type of organisation	e.g. SMEs, NGOs, corporations, government agencies
Type of end users involved	e.g. public institutions, businesses, NGOs, journalists, citizens
Type of service provided	e.g. data visualisation, quality enhancement, end-user outreach
Relation with other intermediaries	e.g. overlapping, complementary
Trust mechanism in place	e.g. certification, reputation, transparency about methods
Impact domains	e.g. biodiversity (environmental), anti-corruption (social), consumer applications (economic), etc.

The results show the sheer variety of cases available, and the different kinds of impact they can have. The next section provides a more detailed overview of the insight gained from the literature review.

### 3.2 What research says about downstream open data reuse

Few robust analyses exist of the role that intermediaries play in open government data value chains. Existing reports tend to examine intermediary activity within the **broader data ecosystem** rather than focus selectively on intermediary use of open government data and/or the value created by the presence of intermediaries in the open data sector.

Recent literature has focused heavily on the potential of data intermediaries to facilitate the sharing and aggregation of private data held by individuals. This is unsurprising given the broader growth in recent years of concerns regarding data privacy and data collection by large corporate entities in the technology sector. While intermediaries have the potential to improve consumer trust in data sharing transactions by functioning as third-party alternatives to corporate data aggregators, this function falls outside the scope of this report’s analysis.

A second factor complicating the analysis of intermediary use of open government data is the **heterogenous definitions** of ‘intermediary’ or ‘infomediary’ found in existing literature. Many definitions situate intermediaries between data providers and end users, but both portals created by data-publishing administrations and reuse applications fall within these parameters.

Where function-based typologies of intermediaries are provided, they generally bear similarities to the sixfold typology laid out in the study of the role of intermediation in overcoming barriers to open data use by Den Haan (2018). This study categorised intermediaries as ‘aggregators’, ‘communicators’,

‘demanders’, ‘developers’, ‘producers’ and ‘validators’, with a single intermediary sometimes fulfilling multiple roles (even within the same data chain). Here, ‘aggregators’ – intermediaries that compile data from multiple public sources and offer it in repackaged form – most closely align with the definition used in this report. In his chapter on repurposing public data, Loshin (2013) uses ‘aggregator’ and ‘intermediary’ interchangeably to refer to organisations that collect and compile data and provide ‘value-added processing’. Roles proposed elsewhere in the literature that fall within the scope of this report include ‘enablers’, who facilitate the use of data by creating platforms that enable extraction, and ‘enrichers’, who broadly add value to data through aggregation, cleaning, etc. (Berends, Carrara and Radu, 2020).

Some reports define intermediaries by their **function** rather than their position in the value chain. In their study of the role played by intermediaries in open data ecosystems, Van Schalkwyk, Willmers and McNaughton (2016) focus on the ability of intermediaries to facilitate the flow of viscous data and consequently only specify that intermediaries ‘increase the accessibility and utility of data’. Based on this definition, the broadest one found, intermediaries do not even need to directly interact with data; brokering data-related agreements between agents in a value chain constitutes intermediary activity. A smaller number of studies list **specific services** that intermediaries might provide. Janssen and Zuiderwijk (2014) have identified six: ‘single-purpose apps’, ‘interactive apps’, ‘information aggregators’, ‘comparison models’, ‘open data repositories’ and ‘service platforms’. While single-purpose apps, interactive apps and most open data repositories do not fall within the definition used by this report, it is worth noting that in some cases intermediaries may provide these services while simultaneously aggregating and republishing data for reuse. The Johns Hopkins Coronavirus Resource Center, for example, both creates visualisations from COVID-19-related data and allows for cleaned, compiled datasets to be downloaded in CSV format for further reuse <sup>(99)</sup>.

## Intermediary positioning in healthy data ecosystems

The existing literature broadly agrees on the factors that contribute to healthy data-sharing environments. The most important ones are the accessibility and liquidity of data. The World Economic Forum (2022) makes a distinction between the **data ecosystem**, which comprises all data, all data transactions and all spaces in which data exists and is processed, and the **data value chain**. The value chain is a model of the flow of data from publishers to end users developed by Open Data Watch and subsequently adopted broadly in studies of data transactions. It is divided into four stages: collection, publication, uptake and impact. Since data may be reused an infinite number of times and in an infinite number of ways, the chain also includes a feedback loop. The existence of open data in an ecosystem is not synonymous with its entry into the value chain, since opening up data does not inherently guarantee its movement from producers to end users or even from producers to reusers. Van Schalkwyk, Willmers and McNaughton (2016) made reference to ‘viscous data’: data that, once opened, remains inert rather than becoming a stable, dynamic aspect of the data ecosystem.

**Barriers** to reuse vary widely. While the concept of open data is not new, many coordinated data opening initiatives are relatively recent. Governments are still in the process of taking on the role of data producers <sup>(100)</sup>. Johnson and Greene’s survey of government open data intermediary activity in major Canadian cities found that national and municipal administrations across North America tended

---

<sup>(99)</sup> Gao and Janssen, 2022.

<sup>(100)</sup> Johnson and Greene, 2017.

to provide data in specialised formats. Moreover, since data was frequently labelled with government-specific jargon, determining the contents of specific datasets could pose a challenge. In some cases, governments presented aggregated data without providing adequate information regarding its scale and collection process<sup>(101)</sup>. Other studies identified the inconsistent inclusion of metadata as a complicating factor for reuse. Loshin (2013, p. 289) criticised the frequent presentation of open data in ‘arbitrary’ and/or ‘artificially constructed’ formats and noted that datasets were sometimes provided in ‘linearised’ forms that rendered interpretation difficult.

Studies that conceptualised the data environment as an ecosystem, such as Den Haan (2018), frequently identified intermediaries as ‘**keystone species**’, meaning that they facilitate the flow of data by providing services that enable access to other actors, serving as an intermediate link in data chains. Open data intermediaries may enter the data ecosystem to facilitate interactions between agents with specific intents, but unequal capacities. In practice, one agent generally possesses an asset (data) and the other lacks the material resources or capital required to access it, necessitating third-party intervention<sup>(102)</sup>. Data may flow through multiple intermediaries before reaching a reuser. Alternatively, intermediaries may specialise in more than one form of data processing, providing services at multiple points along the same supply chain<sup>(103)</sup>. In their study of open data intermediaries in developing countries, Van Schalkwyk et al. (2015, p. 15) interpreted the presence of multiple intermediaries in open data supply chains as a positive indicator for data liquidity insofar as it suggested greater potential for reuse. That said, it is necessary to verify that intermediaries are serving in complementary, non-redundant capacities. Similarly, although the existence of large numbers of intermediaries in the open data ecosystem may be interpreted as a positive signifier of intermediary specialisation and intermediary-enabled access, it is a weaker indicator for impact than the actual value created by intermediary activity.

As illustrated by the case of intermediaries in South-African academia (see Box 1), they may operate **concurrently** in a very similar capacity and still independently improve the data ecosystem. While intermediaries are third parties in data transactions, they exercise influence over data, either by selectively facilitating certain interactions or by directly altering the type and scope of data made available. Intermediaries operate within a ‘field’, in which established actors attempt to preserve existing structures (e.g. data chains) while new actors attempt to disrupt them<sup>(104)</sup>. Disruption in this sense does not mean a severing of the link between the data source and the reuser. It simply suggests alterations to the chain to incorporate new intermediaries, intermediary services, etc.

---

<sup>(101)</sup> Johnson and Greene, 2017, p. 10.

<sup>(102)</sup> Van Schalkwyk et al., 2015, p. 10.

<sup>(103)</sup> Van Schalkwyk et al., 2015, p. 48.

<sup>(104)</sup> Van Schalkwyk et al., 2015, p. 9.

**Box 1**

Van Schalkwyk, Willmers and McNaughton (2016) discuss a South African case study where two separate intermediaries facilitate access to open data on the governance of public universities. This information is submitted by universities to the national government and stored in the South African Department of Higher Education and Training's higher education management information system (HEMIS), an SQL database.

Both the Centre for Higher Education Trust (CHET), a non-governmental organisation, and IDSC, a private IT company, act in intermediary capacities to supply extracted and organised data to universities for comparative analysis. These organisations do not coordinate or cooperate; it is entirely possible that there is overlap in the services that they provide.

The existence of multiple intermediaries in interactions between HEMIS and universities likely helped to mitigate the 'norms and values' embedded in published data. CHET's datasets, for example, are organised according to indicators, which may (intentionally or unintentionally) reflect institutional views or biases.

## Intermediary-facilitated access to data

Intermediaries are able to **leverage various types of capital** to facilitate access to open data. By employing technical, social and cultural capital, they carve out niches in data value chains and compensate for deficits in both producer and reuser capacities.

**Technical capital** is a commonly identified asset, referenced directly or indirectly in six studies. Public administrations that publish open data frequently lack the technical infrastructure to provide data in homogenous, easily accessible formats. Consequently, intermediaries are able to repackage, reformat and republish this data in forms that have **greater reuse potential** <sup>(105)</sup>. In their study of open data intermediaries in developing countries, Van Schalkwyk et al. (2015, p. 13) found that in 97 % of analysed cases, these intermediaries deployed technical capital to 'collect, digitise, clean, reorganise and translate data'. According to De Haan (2018), intermediaries also provide more 'flexible' data models, allowing for more efficient data linkage. Den Haan (2018) identified two cases in which intermediaries working with geospatial data provided access to cleaned and organised datasets through their own servers, and one in which an intermediary designed mechanisms to automatically update datasets when administrations released new data. Other intermediaries created superior mechanisms for finding republished data, including 'smart' keyword-based searching. Here, the intermediary was able to function as a sector-specific alternative to the national open data portal. The deployment of technical capital in isolation, however, while sufficient to enable transactions, does not always adequately consider the needs of citizens. Consequently, technical capital is commonly deployed in conjunction with social, economic, cultural and symbolic capital. Van Schalkwyk et al. (2015, p. 17) posit that technical expertise and capacities could be proxies for other established types of asset. Intermediaries may, for example, be able to acquire technical capacities by making use of pre-existing economic capital to purchase information-technology infrastructure.

**Social and cultural capital** are indirectly or directly referenced in six and five studies respectively. Intermediaries, data producers and data itself all possess social and cultural capital to various extents. Public administrations may lack the requisite social capital to recognise the potential value of data

<sup>(105)</sup> Berends, Carrara and Radu, 2020, p. 16.

reuse and may consequently provide data in inaccessible or incomplete forms, regardless of their technological capacities<sup>(106)</sup>. When intermediaries leverage their technical capital to facilitate the reuse of data by non-technical actors, they **improve public perception of the quality** of open data. This has the potential to increase the cultural and social capital of the provided data, increasing public trust in data providers and in the data ecosystem as a whole. Intermediaries may also leverage their pre-existing social and cultural capital to increase the reach of open data<sup>(107)</sup>. The Johns Hopkins Coronavirus Resource Center's COVID-19 data was widely reused and analysed, not only because of its inherent value, but also because of the institution's prestige and existing reputation for scientific excellence<sup>(108)</sup>. Johnson and Greene (2017) classified media organisations that produced articles and visualisations based on data as intermediaries. Although this falls outside the scope of this report, their conclusion that media intermediaries possess influence disproportionate to their comparatively small presence in the intermediary sector due to their wide readership and production of public-facing products reinforces this idea. Open data intermediaries often specialise in particular types of data. As a result, over time, they may **accumulate social capital in specific sectors**. Unsurprisingly, Van Schalkwyk et al. (2015, p. 14) found that in cases where intermediaries had reputations for producing high-quality data, clients had higher confidence in them and were more likely to retain their services. Recognising the value of this capital, intermediaries generally develop and employ **strategies to 'identify, attract and retain'** specific customer bases or groups of users<sup>(109)</sup>.

Where mechanisms exist for communication between open government data providers (e.g. public administrations) and intermediaries, intermediaries may also use their cultural capital to **lobby** for further opening of data and improvements to existing open datasets. Crusoe and Melin (2018) note in their literature review of barriers to the use of open government data that increasing the cultural capital of open data may also increase public demand for continued publishing. The potential downside, however, is that data providers may become dependent on the services provided by open data intermediaries and consequently feel no motivation to open data in such a way that it may be directly processed by users<sup>(110)</sup>. Studies of intermediary activity found that most intermediaries were private-sector companies. We can assume that these intermediaries act to maximise profits and to ensure their ability to continue to operate in the data sector<sup>(111)</sup>. As a result, intermediaries may not be likely to lobby for the removal of barriers to immediate use. In developing countries, a significant majority of studied intermediaries (72 %) were non-profit organisations reliant on donor funding, so this issue may be more relevant in developed countries with larger private intermediary sectors<sup>(112)</sup>.

In their study of network relations in open government data ecosystems, Reggi and Dawes (2022, p. 5) found that connections between intermediaries and governments are more likely to develop and to persist when these actors are operating in **geographic proximity**. This impact was particularly strong at the municipal level. This can likely be attributed to a combination of practical feasibility and cultural capital. Through repeated interactions, intermediaries may also forge closer relationships with administrations. Van Schalkwyk et al. (2015, p. 14.) cited the example of an intermediary in South Africa that is able to consistently facilitate access to open data on public higher education due to its

---

<sup>(106)</sup> Loshin, 2013, p. 289.

<sup>(107)</sup> Johnson and Greene, 2017, p. 14.

<sup>(108)</sup> Gao and Janssen, 2022, p. 8.

<sup>(109)</sup> Janssen and Zuiderwijk, 2014, p. 698.

<sup>(110)</sup> Janssen and Zuiderwijk, 2014, p. 706.

<sup>(111)</sup> Johnson and Greene, 2017, p. 11.

<sup>(112)</sup> Van Schalkwyk et al., 2015, p. 18.



'long-standing' relationship with personnel in the government department that aggregates this data. Finally, in cases where data is partially open (e.g. when right-to-information legislation mandates that government data be released upon request, but data cannot be downloaded from an easily accessible portal), intermediaries may deploy their cultural capital and/or connections to administrations to access this data and render it more available for public consumption <sup>(113)</sup>.

## The social, political, economic and environmental impact of intermediaries

### Social impact of open data intermediaries

Open government data has the potential to be applied and reused in service of social justice. Thus, by increasing its accessibility and reuse potential, intermediaries may indirectly create social impact. Multiple assessments of intermediary impact specifically referenced the utility of visualisations in communicating data-based insights to the public <sup>(114)</sup>. In addition to being more comprehensible for the layperson, visualisations are well suited for dissemination across social media networks <sup>(115)</sup>. During the pandemic, visualisations using COVID-19-related open data published by governments were widely shared on social media, but many of these either cherry-picked data or represented it incorrectly <sup>(116)</sup>. Intermediaries facilitate the reuse of data towards this end, but are likely to have superior capacities to interpret data compared to laypersons. As a result, when intermediaries are present in this sort of data transaction, the likelihood of **misinterpretation or misrepresentation can decrease**. In developing countries, where social, economic, cultural and technological capital are more likely to be unevenly distributed, capital-rich intermediaries serve an even more critical role in **democratising** access to data <sup>(117)</sup>. Moreover, since the entry of laypersons into the data environment increases demand for intermediary services, intermediaries have a vested interest in encouraging **public interest** in data. While a majority of studies observed intermediaries acting indirectly to stimulate this interest, Den Haan (2018, p. 43) interviewed one intermediary company that hosted events biannually to facilitate in-person meetings between data producers and data reusers. Similarly, Van Schalkwyk et al. (2015, p. 32) identified an Indian energy sector intermediary that periodically organised conferences to discuss energy-related issues. This suggests that the impact of intermediaries may extend beyond their direct interaction with data. The strategies that they employ to attract customers may also influence the nature of data transactions.

By conducting **market research**, intermediaries may be able to better assess public demand for specific types of data and data-related services than data providers. Johnson and Greene (2017, p. 15) make reference to cases where intermediaries have applied data in contexts that were likely not anticipated by providers. The specific case that they cite is a point-and-click game based on neighbourhood boundary data provided to enable mapmaking. While this sort of application development falls outside this report's definition of intermediation, intermediaries that aggregate and enrich data may similarly fill previously unidentified gaps in open data markets. By leveraging their social and cultural capital, intermediaries may also be able to create **higher-trust research environments** to facilitate research in the service of public interest <sup>(118)</sup>.

<sup>(113)</sup> Van Schalkwyk et al., 2015, p. 14.

<sup>(114)</sup> Den Haan, 2018, p. 30.

<sup>(115)</sup> Janssen and Zuiderwijk, 2014, p. 695.

<sup>(116)</sup> Ackerman, 2021.

<sup>(117)</sup> Van Schalkwyk, et al., 2015, p. 8.

<sup>(118)</sup> Centre for Data Ethics and Innovation, 2021, p. 14.

A **common concern** raised in existing literature regarding intermediary handling of data is that the for-profit nature of many intermediaries will **disincentivise** adherence to ‘voluntary fiduciary duties of care and loyalty’, particularly in the absence of comprehensive regulations governing the use and transfer of data <sup>(119)</sup>. While this concern is largely inapplicable to open data intermediaries because the data that they use is not private, intermediaries that charge for their services cannot contribute to the democratisation of data access to the same extent as their not-for-profit peers. In his post on the OpenFuture blog on the Data Governance Act, Vogelezang (2022) criticised the market structure recommended by the European Commission on the grounds that it will not lead to a meaningful shift away from ‘informational capitalism’; data sharing will still be dependent on platformised business models and intermediaries will be incentivised to centralise their power. Non-profit intermediaries, however, face their own set of operational challenges – they must be sufficiently economically **viable** to cover the ongoing costs of data storage and processing and their autonomy may be limited by dependence on third-party funding <sup>(120)</sup>. Vogelezang (2022) noted that critics of the Data Governance Act have questioned the viability of establishing data intermediation services independent from ‘for-profit activity’, anticipating that these intermediaries will find it challenging to sustain themselves in the absence of means for data monetisation.

Van Schalkwyk et al. (2015, p. 18) posited that in some cases, intermediaries might **fabricate user deficits** to justify their existence in open data ecosystems, particularly in low-trust environments or environments with highly disparate capital distributions where reusers are highly dependent on intermediaries for access. While their study found no evidence of this phenomenon in the developing countries surveyed, it may still be necessary to monitor intermediary activity to ensure that barriers to access are not presumed to exist or created. A number of articles and papers published in response to the Data Governance Act, however, noted the **difficulties associated with establishing guidelines** for intermediary activity. In their paper on altering the direction of the European data ecosystem, Chapman et al. (2021) proposed that a government scheme be created to identify responsible data intermediaries, but acknowledged that a necessary first step would be the creation of a comprehensive definition of responsible data sharing.

### Economic impact of open data intermediaries

The entry of new intermediaries into the open data ecosystem is a positive trend insofar as it suggests **growth, innovation and competition** within the sector. ASEDIE’s 2021 survey of infomediary companies in Spain, one of a very small number of quantitative assessments of the impact of data intermediaries, calculated the sector’s annual growth, in part by tracking the creation of these companies. ASEDIE (2021, p. 5) provides a number of indicators related to the growth and profit of the intermediary sector, including the average age, number of employees and profit of infomediary corporations. ASEDIE also calculated aggregated turnover in the sector and average profit in each of 10 subsectors (e.g. tourism) <sup>(121)</sup>. In their survey, Johnson and Greene (2017, p. 11) similarly assessed the distribution of intermediaries across subsectors, although their typology was based on organisation type rather than the category of data used. In order of prevalence, intermediaries came

---

<sup>(119)</sup> World Economic Forum, 2022.

<sup>(120)</sup> World Economic Forum, 2022.

<sup>(121)</sup> ASEDIE, 2021, p. 19

from the private sector, government, media, non-governmental organisations and academic institutions.

ASEDIE's analysis was primarily retrospective rather than forward-facing, except for the fact that they surveyed the extent to which representatives from infomediary companies anticipate continued 'positive evolution' in the sector if data continues to be opened <sup>(122)</sup>. ASEDIE also used extrapolation based on previous growth trends to estimate sales and employee data for companies that cannot provide data for the most recent year <sup>(123)</sup>.

Besides the economic impact in terms of a growing sector of intermediary businesses, there is also the impact on the businesses that are reusers of the data provided by intermediaries. A majority of existing literature on the relationship between intermediaries and private-sector enterprises focuses on the potential ability of intermediaries to serve as **trusted third parties** for the collection and aggregation of personal and demographic data. Two studies, however, at least briefly addressed the potential for additional value to be created through the intermediary-facilitated **sharing of public data** with these enterprises. In its report on advancing digital agency through intermediation, the World Economic Forum (2022, p. 30) estimated that the ability of private organisations to combine private data with data from open sources has led to a 5–6 % increase in productivity as a result of more effective decision-making. This kind of combination is common in the private sector – ASEDIE found that 70 % of infomediary companies used both open and private data, while only 30 % exclusively used data from public sources <sup>(124)</sup>. An analytical report on the economic benefits of open data also noted that open data was frequently used as a complement to other services (for example, to contextualise private data) <sup>(125)</sup>. In some cases, intermediaries have also worked cooperatively with private-sector enterprises to supplement their technical capabilities. During the early stages of the pandemic, the Johns Hopkins Coronavirus Resource Center partnered with Amazon, Slack and Github to access IT resources required to process COVID-19-related data <sup>(126)</sup>, and the South African Department of Higher Education and Training contracts private IT companies to maintain the SQL database in which their aggregated data is stored <sup>(127)</sup>.

### Political impact of open data intermediaries

Multiple works made reference to situations in which public administrations rely on open data intermediaries to process data. Ackerman (2021) pointed to the pandemic-era reliance of administrations on intermediaries to interpret data and to consolidate data from national governments. In the case study on South Africa presented by Van Schalkwyk, Willmers and McNaughton (2016), the dataset compiled by CHET from raw data compiled by the Department of Higher Education and Training was used by university planners due to its superior formatting when compared to the Department of Higher Education and Training's published dataset. According to Den Haan (2018, p. 10), where **governments recognise the value of reuse**, they may directly fund and in some cases operate intermediary platforms. In the Netherlands, the government has contracted private intermediaries to develop applications for public use. Other national governments have

---

<sup>(122)</sup> ASEDIE, 2021, p. 25.

<sup>(123)</sup> ASEDIE, 2021, p. 5.

<sup>(124)</sup> ASEDIE, 2021, p. 24.

<sup>(125)</sup> Berends, Carrara and Radu, 2020, p. 13.

<sup>(126)</sup> Gao and Janssen, 2022, p. 10.

<sup>(127)</sup> Van Schalkwyk, Willmers and McNaughton, 2016.

employed various schemes to encourage use of open data by non-governmental organisations, private-sector enterprises, academic researchers and individuals. In many cases, however, these initiatives have focused on the direct use of open data by reusers without employment of intermediary services (although given the disagreements regarding the definition of data intermediaries, this exclusion may not reflect a conscious desire to eliminate intermediaries from data value chains).

Intermediaries also have the capacity to **help administrations save money** by taking on costs related to the development and maintenance of IT-related infrastructure. In cases where government departments face resource shortages or institutional constraints, intermediaries may step in to provide necessary capital and flexibility <sup>(128)</sup>. Intermediaries also frequently **take on responsibilities** that previously fell on data providers, not only in regard to enabling access to data, but also to providing technical assistance to reusers. Den Haan noted that many users of Actueel Hoogtebestand Nederland, a government-provided digital height map of the Netherlands, reported that receiving user support third-party intermediaries <sup>(129)</sup>. If intermediaries fail to achieve necessary levels of public visibility, however, these positive impacts may not manifest. Despite the presence of two separate intermediaries collecting and processing data from HEMIS, the Department of Higher Education and Training reported that it was still frequently contacted by would-be reusers asking the department to directly provide them with data <sup>(130)</sup>.

Intermediaries may also **improve citizen perceptions of governments**. When trust was discussed in the context of intermediary-facilitated data transactions, it frequently referred to trust between citizens and corporate entities – situations in which individuals are unwilling to release their personal data to corporations, but may be more willing to disclose this information to third parties. A smaller number of studies examined the potential of intermediaries to increase the cultural capital of data and, as a result, create greater trust between the public and data-publishing governments. One limitation, however, is that in cases where intermediary interpretations of data are found to be flawed, citizen trust in administrations may decrease. McLean et al. (2021) examined the repercussions of a case in which Spanish Prime Minister Pedro Sánchez cited statistics regarding Spain’s COVID-19 testing volumes in an official briefing; these statistics, which came from a third-party data aggregator and visualiser, were later discovered to be incorrect. Subsequently, critics raised questions as to why the government had relied on this intermediary given that its methodology was opaque and its expertise was largely unverified. In order to maintain credibility, governments must exercise discretion with regard to the intermediaries whose services they employ.

Finally, intermediaries may be able to **provide governments with feedback** on their data-opening processes. On the one hand, intermediaries may be able to lobby for greater interoperability in the data published by administrations (for example, common formats and enforced quality standards) <sup>(131)</sup>. On the other hand, when intermediaries conduct market research to determine demand for data-related services, they may identify sectors where discrepancies exist between what is supplied by producers and what users require <sup>(132)</sup>. If intermediaries are given the opportunity to report these

---

<sup>(128)</sup> Van Schalkwyk, Willmers and McNaughton, 2016.

<sup>(129)</sup> Den Haan, 2018, p. 10

<sup>(130)</sup> Van Schalkwyk, Willmers and McNaughton, 2016.

<sup>(131)</sup> Centre for Data Ethics and Innovation, 2021, p. 15.

<sup>(132)</sup> Den Haan, 2018, p. 22.

disconnects, governments may be able to identify high-demand datasets and prioritise their opening accordingly.

### Environmental impact of open data intermediaries

The existing literature tends to focus on the social and political impacts of open data intermediaries. Data producers and end users were the most frequently identified beneficiaries. A smaller number of studies, such as ASEDIE (2021), examined the broader economic impact of growth in intermediary sectors. The environmental impact of intermediary-facilitated reuse is **not explored in detail** in any of the reviewed studies. While some of the open data intermediaries profiled by Van Schalkwyk et al. (2015, p. 32) operate in relevant fields, the authors do not specifically examine the potential environmental benefits. The company Indianpetro, for example, aggregates and republishes data from the Indian energy sector, which in theory may be used by academics, industry professionals and lobbyists in service of environmental goals.

## 3.3 Conclusion: considerations for the development of an open-data impact-assessment methodology

There is no robust aggregate analysis of the importance of data reuse through intermediaries. However, available evidence shows that the role of intermediaries is considerable – in some cases it is greater than that of the original open data publisher. When assessing the impact of open data, it is therefore imperative to consider the indirect impact generated through intermediaries. Otherwise, the actual impact might be under-estimated. The available literature does not attempt to calculate the possible size of this underestimation.

Depending on the chosen impact assessment methodology, measuring the impact of open data intermediaries is methodologically similar to measuring the impact of open data through direct data reuse. As shared in Section 1 (where open data impact is measured based on the economic, environmental, social and political domains) and Section 2 (regarding open data impact indicators on portals, such as the number of downloads), there are pitfalls in measuring the impact of open data intermediaries because dimensions are inconsistent or arbitrarily chosen and the tools selected and used cannot accurately measure impact. Thus, taking these difficulties into account, this section presents how the impact of intermediaries could be captured in the context of an **existing impact assessment methodology**, rather than by developing a set of self-standing indicators. In other words, the goal is to measure the difference that open data intermediaries make, with respect to traditional frameworks for impact measurement of open data.

The measurement framework used in this context is the framework used for the ODM report. Table 5 shows how the role of intermediaries can be tracked and measured across the different dimensions of the framework. Based on the above literature review, indicators and sources for capturing the role of intermediaries within existing frameworks are proposed. The indicators presented here have different levels of feasibility and relevance, depending on the methods used. Web surveying implies a systematic analysis of the websites of the intermediaries and can be carried out at relatively low costs. Some features, such as usability, entail expert assessments, which are often costly and can be subjective.

Broadly speaking, indicators that can be detected through web surveying – mainly those related to the dimensions of open data portal and quality – have higher feasibility. Impact indicators require ad hoc primary data collection, including some forms of surveying, and are therefore very important but also more challenging. The sweet spot might lie in indicators related to usage, which can be detected through web surveying, if data are published by the intermediary.

Measuring social, political, economic and environmental impact requires desk research and surveys of the use case owners, and would therefore be time consuming. The number of additional use cases of open data facilitated by the open data intermediary might be measurable with web surveys or forms of automated tooling, but to know the actual impact of a use case requires careful research and estimation.

Overall, as was also highlighted in Section 3, the most important methodological choice lies in **standardisation**: reusing standard indicators and sources as much as possible, to ensure that the indicators collected are comparable across different intermediaries.

Table 5: Measuring the impact of open data through intermediaries

Open Data Maturity Framework	Expected impact of open data intermediary	Possible indicators	Possible source	
Open Data Portal	Portal features	Increased usability and user friendliness	Difference with main open data publisher in terms of usability score and availability of additional features	expert assessment
	Portal usage	Increased usage ; potential new users	number of monthly active users; number of downloads or API calls; percentage of users not having used original data sources or having switched to the open data intermediary; percentage of access/download to open data portal coming from open data intermediaries	survey of intermediaries survey of (re)users
	Data provision	Data from multiple sources, not only governments	Difference with open data published in terms of number of datasets provided by data intermediary; number of different sources for datasets; number of topics within the scope of the intermediary	web survey
	Portal sustainability	Much more important and challenging for new players; possible reducing cost for open data publishers	number of open data intermediaries mapped; availability of competitive funding from private and public sector	web survey survey of intermediaries
Open Data Quality	Currency	Depending on the original open data, hence cannot be better than original	Comparative delay with original data	web survey
	Monitoring and measures	More complex monitoring because of decentralised data delivery	Difference in transparency score for the reporting of the open data intermediary, e.g. publication of monthly active users, download statistics	web survey
	DCAT-AP compliance	Higher or lower compliance compared to original data	Difference in compliance score with open data publisher	expert assessment
	Deployment quality and linked data	Higher or lower quality	Difference in quality score with open data publisher	expert assessment
Open Data Impact	Strategic awareness	Same as monitoring	Same as monitoring	Same as monitoring
	Political impact	Greater transparency, accountability, and democratic participation	Number of additional reuse case for political impact (apps, online services, citations) Difference in number of reuse case with open data publisher Concrete impact generated by reusers of open data intermediaries	desk research survey of reusers
	Social impact	Increased participation of marginalised groups	Number of additional reuse case for social impact (apps, online services, citations) Difference in number of reuse case with open data publisher Concrete impact generated by reusers of open data intermediaries	desk research survey of reusers
	Economic impact	new innovative solutions; business and job creation; new applications for addressing societal challenges	Number of additional reuse case for economic impact (apps, online services, citations) Difference in number of reuse case with open data publisher Revenues generated by reusers of open data intermediaries	desk research survey of reusers
	Environmental impact	innovative solutions to address climate change; increased awareness of climate change	Number of additional reuse case for environmental impact (apps, online services, citations) Difference in number of reuse case with open data publisher Concrete impact generated by reusers of open data intermediaries	desk research survey of reusers

## 4 Impact indicators that can be obtained via automated feedback mechanisms

As a final contribution to the report on the impact of open data, this section explores whether any of the indicators and metrics that have been proposed in existing indicator frameworks can be generated computationally through **automated mechanisms** (e.g. data science methods over existing published data, dataset citation in the literature, API call logs for dynamic data). It also explores whether there is any opportunity to add additional indicators to existing indicator frameworks, as long as these can be generated automatically.

Section 4.1 provides the context and justification for the work presented here, setting the basis of why automation is relevant to facilitate impact assessment. Section 4.2 summarises the main findings from previous analyses performed in the context of data.europa.eu, as well their recommendations to automate the calculation of some of the identified impact metrics. Subsequently, Section 4.3 gives a systematic literature review with the aim of identifying recent works that may have proposed techniques and tools for the automation of impact calculation, or new metrics that can be obtained via automation. Finally, Section 4.4 concludes with recommendations on next steps.

### 4.1 Context: automation and open data impact

As previous research from data.europa.eu on the sustainability of open data ecosystems <sup>(133)</sup> describes, there is a strong relationship between impact and use and reuse of data, though the two cannot be used interchangeably. One of the definitions surveyed in earlier sections (see Table 1) captures the essence of this relationship: Corrales-Garay, Ortiz de Urbina Criado and Mora-Valentín (2020) state that impact ‘addresses the effects of reusing open data and the innovation that has been created’.

The impact methodologies discussed in the previous sections cover a range of impacts, reported in the form of indicators at the macro- and micro-level. While there are various approaches to defining and assessing open data impact, our ability to automate them depends on the format in which the impact data can be made available and on the technologies and tools that can be used to track such data.

The possibilities of automating the measurement of open data impact in the social, political, economic and environmental domain as described in Section 1, are limited. For instance, the ODM collects data via questionnaires, surveys and interviews; from a technical point of view, the data is a mix of numerical, categorical and free text data, which to a large degree cannot be collected automatically as it requires access to the knowledge and expertise of the publishers. However, automation can be useful in measuring the impact of open data portals, by using indicators such as those described in Section 2. Automation can be used to compile core metrics such as the number of datasets, API requests, downloads, common search queries, registered users, etc. Such statistics may be used by both open data portals and respondents to the ODM report. Such metrics can be computed if the portal where the open government datasets are published is equipped with the relevant capabilities.

---

<sup>(133)</sup> Publications Office of the European Union, 2020.



Some open data impact methodologies consider microeconomic indicators, which could be linked to proxies that may be obtained via automatic mechanisms: for instance, the annual ASEDIE infomediary sector reports track value creation in public and private sector institutions, including the progress over time of data-driven businesses, using a series of standard metrics to evidence economic value.

In addition, data science methods could be used, where relevant, to process the data at scale, for example to run sentiment analysis on the responses or to identify key entities such as countries, organisations, topics, etc. through the use of techniques such as topic modelling, named entity recognition and geotagging. For instance, running named entity recognition techniques over the texts obtained via the questionnaires run for the ODM report could allow to identify the names of active departments and national agencies identified in those reports, along with the names of and references to software packages and tools that are mentioned in the reports as examples for open data reuse.

## 4.2 Starting point from previous research on European data portals and an initial set of proposals for automating the calculation of use and reuse indicators

As noted in Sections 2 and 3 of this report, a link between portal activity and actual use or even impact remains challenging. Walker, Frank and Thompson (2015) suggest six criteria for ‘good metrics’ of open data: desirable metrics should be valid (closely correlated with the attribute of interest), reliable (consistent results over time and between observations), discriminatory (or sensitive to common values of the attribute of interest), efficient in its assessment and use, transferable to different publishing and sector contexts, and comparable to establish meaningful baselines and study how the metrics evolve. An ideal metric would rate highly on all of these criteria, and would allow for an efficient assessment (e.g. automated) to be quickly run against a large group of datasets with high validity and reliability, yielding results that are comparable for a wide range of contexts.

Previous research of the Publication office of the European Union (2020) proposed a series of operational user-centric metrics, organised according to 10 dimensions. These metrics can be used by portal owners to assess the current sustainability of their portal. The metrics are summarised in Table 6 and were tested on 10 different portals with varying levels of open data maturity.

*Table 6: User-centric metrics and how they can be operationalised in existing data publishing portals, from Publications Office of the European Union (2020).*

Dimension	Metric	Existing/developed
<b>Organise for use</b>	<ol style="list-style-type: none"> <li>1. Each dataset is accompanied by a comprehensive descriptive record (going beyond a collection of structured metadata).</li> <li>2. An extract of the data can be previewed (for easier sense-making).</li> <li>3. The portal provides recommendations for related datasets.</li> <li>4. The portal enables users to review/rate the datasets.</li> <li>5. Keywords from datasets are linked to other</li> </ol>	Based on Opquast’s web data quality checklist  <a href="http://opquast.com/en/">http://opquast.com/en/</a>

	published datasets	
<b>Co-locate documentation</b>	<ol style="list-style-type: none"> <li>1. Supporting documentation does not exist.</li> <li>2. Supporting documentation exists but as a document which has to be found separately from the data.</li> <li>3. Supporting documentation is found at the same time as the data (e.g. the link to the document is next to the link to the data in the search).</li> <li>4. Supporting documentation can be immediately accessed from within the dataset but it is not context sensitive (e.g. a link to the documentation or text contained within the dataset).</li> <li>5. Supporting documentation can be immediately accessed from within the dataset. It is context sensitive so that users can immediately access information about a specific item of concern (e.g. a link to a specific point in the documentation or the text contained within the dataset).</li> </ol>	Intelligibility metric (Walker, Frank and Thompson, 2015)
<b>Be measurable</b>	<ol style="list-style-type: none"> <li>1. Portal has no analytics.</li> <li>2. Portal has site analytics.</li> <li>3. Portal has use analytics.</li> <li>4. Portal has impact analytics.</li> </ol>	Based on review of web analytics tools
<b>Promote standards</b>	<ol style="list-style-type: none"> <li>1. A permanent, patterned and/or discoverable URI/URL is used for each dataset (URI/URLs can be used as universal, unique identifiers by appending a serial number or other internal naming system to a domain).</li> <li>2. The portal uses versioning of datasets (to maintain the history of a dataset).</li> <li>3. Dates are available in a standard format (facilitates the automated exploitation of date-type data and their conversion according to specific needs or constraints).</li> <li>4. Metadata associated with each dataset is available in a standard format (e.g. using VOID or DCAT) to enable automated metadata retrieval and the import of metadata from other data catalogues.</li> <li>5. The metadata catalogue can be retrieved using a standard protocol (e.g. automatic retrieval of the metadata catalogue using RDF or HTTP GET).</li> </ol>	Based on guidelines from the World-Wide Web Consortium's eGov Interest Group and OpQuast.
<b>Promote metadata</b>	<ul style="list-style-type: none"> <li>★ Metadata ignorance.</li> <li>★★ Scattered or closed metadata.</li> <li>★★★ Open metadata for humans.</li> <li>★★★★ Open reusable metadata.</li> <li>★★★★★ Linked open metadata.</li> </ul>	European Commission 5-level maturity schema for metadata management

<p data-bbox="229 226 483 707"><b>Link data</b></p>	<p data-bbox="483 226 1174 315">★ <b>On the Web.</b> Make your stuff available on the Web (whatever format) under an open license.</p> <p data-bbox="483 315 1174 427">★★ <b>Machine-readable data.</b> Make it available as structured data (e.g. Excel instead of image scan of a table).</p> <p data-bbox="483 427 1174 517">★★★ <b>Non-proprietary format.</b> Make it available in a non-proprietary open format (e.g. CSV instead of Excel).</p> <p data-bbox="483 517 1174 607">★★★★ <b>RDF standards.</b> Use URIs to denote things, so that people can point at your stuff.</p> <p data-bbox="483 607 1174 707">★★★★★ <b>Linked RDF.</b> Link your data to other data to provide context.</p>	<p data-bbox="1174 226 1457 707">5 Stars of Open Data, by Tim Berners Lee</p>
<p data-bbox="229 707 483 1211"><b>Promote use</b></p>	<ol data-bbox="483 707 1174 1211" style="list-style-type: none"> <li>1. The portal is connected with social media to create a social distribution channel for open data.</li> <li>2. The portal provides users with online support for feedback, to request/suggest the publication of new datasets, and when problems arise during use (e.g. contact form, discussion forum, FAQs, helpdesk, search tips, tutorials, demos).</li> <li>3. The portal provides a way for users to keep informed of updates to the data (e.g. news feed).</li> <li>4. Datasets are accompanied by links or resources that provide user guidance and support.</li> <li>5. Examples of reuse (fictitious or real) are provided (e.g. information contributed by other users, last reuse, best reuse, data stories).</li> </ol>	<p data-bbox="1174 707 1457 1211">Based on range of literature</p>
<p data-bbox="229 1211 483 1581"><b>Be discoverable</b></p>	<ol data-bbox="483 1211 1174 1581" style="list-style-type: none"> <li>1. The publisher/owner of the data has an open data portal (or similar search mechanism).</li> <li>2. The publisher/owner of that portal publishes an updated, searchable list of datasets.</li> <li>3. The publisher/owner of that portal publishes an updated, searchable list of datasets with synonyms.</li> <li>4. The publisher/owner of that portal publishes a list of datasets which are known to exist but are not currently available (limiting the time wasted on abortive searches).</li> </ol>	<p data-bbox="1174 1211 1457 1581">Discoverability metric (Walker, Frank and Thompson, 2015)</p>
<p data-bbox="229 1581 483 2049"><b>Co-locate tools</b></p>	<ol data-bbox="483 1581 1174 2049" style="list-style-type: none"> <li>1. The portal does not provide visualisation or collaboration tools for users to engage with the datasets.</li> <li>2. The portal provides visualisation tools to enable users to engage with the datasets.</li> <li>3. The portal provides visualisation and collaboration tools to enable users to participate in the governance of the portal (e.g. dataset rating) but the engagement with other users is limited or mediated by the administrator.</li> <li>4. The portal provides visualisation and collaboration tools to enable users to collaborate innovatively with other users.</li> </ol>	<p data-bbox="1174 1581 1457 2049">Based on range of literature</p>

<p data-bbox="229 228 481 598"><b>Be accessible</b></p>	<ol data-bbox="481 228 1174 598" style="list-style-type: none"> <li data-bbox="481 228 1174 331">1. The portal uses human and machine-readable and non-proprietary formats (e.g. CSV, XML, RDF-based formats).</li> <li data-bbox="481 331 1174 398">2. The portal provides different types of formats for the same dataset.</li> <li data-bbox="481 398 1174 465">3. The mechanisms for accessing and interacting with datasets are documented.</li> <li data-bbox="481 465 1174 499">4. Multilingual support is available on the portal.</li> <li data-bbox="481 499 1174 598">5. The portal supports the visually and hearing impaired.</li> </ol>	<p data-bbox="1174 228 1445 398">Based on Web Content Accessibility Guidelines, Version 2.0, from the World-Wide Web Consortium</p> <p data-bbox="1174 432 1445 598">Uses a ‘always, sometimes, never’ scale rather than a cumulative one.</p>
---	---	--

Starting from this operational framework, Dix (2019) proposed specific ways to calculate these metrics for portals using the open-source software CKAN. Table 7 gives an example of the formulas proposed for a category of metrics for promoting reuse, while Table 8 does the same for accessibility metrics.

Table 7: Proposal of formulas to calculate the metrics around promoting reuse, from Dix (2019)..

Promote use: “promote use of open data portals, through the sharing of knowledge and co-opting methods” (Simperl and Walker, 2019, p9):

- The maximum possible score for metrics 2a, 2b, 2d is ‘1’ i.e. 100% of either datasets or webpages;
- The maximum possible score for metric 2c is also ‘1’ i.e. ‘newsfeed present’=1 or ‘newsfeed absent’=0;
- The maximum mean score for this category is therefore also ‘1’.

Metric (Simperl and Walker, 2019)	Natural language description (Simperl and Walker, 2019)	Calculation <sup>1,2,3</sup>	Notation <sup>4</sup>	Source
a. Social media links	The portal is connected with social media to create a social distribution channel for open data	sum webpages with social share code installed / sum webpages	$\frac{\sum_{social=1}^n social}{\sum_{pages=1}^n pages}$ <i>social</i> : Webpages with social share code installed <i>pages</i> : Webpages crawled	Primary source: Simperl and Walker (2019) Secondary validation: Alexopoulos et al. (2014); O’Reilly (2007); Sayogo et al. (2014)
b. Feedback and support	The portal provides users with online support for feedback, to request/suggest the publication of new datasets, and when problems arise during use	sum datasets with at least one completed CKAN ‘contact’ details metadata field / sum datasets	$\frac{\sum_{contacts=1}^N contacts}{\sum_{datasets=1}^N datasets}$ <i>contacts</i> : Datasets with at least one completed CKAN ‘contact’ metadata field <i>datasets</i> : Total portal datasets	Primary source: Simperl and Walker (2019) Secondary validation: Olsina (1999 in Bauer and Scharl, 2000); Huang and Benyoucef (2013); Susser and Ariga (2006); Jaiswal et al. (2010); Liu and Arnett (2000); Flavian et al. (2006)
c. Newsfeed	The portal provides a way for users to keep informed of updates to the data (e.g. news feed)	portal provides newsfeed functionality (1/0)	$N_{feedURL \geq 1} \Rightarrow newsfeed = 1$ $N_{feedURL = 0} \Rightarrow newsfeed = 0$ <i>feedURL</i> : RSS or ATOM newsfeed URL <i>newsfeed</i> : Newsfeed metric score	Primary source: Simperl and Walker (2019) Secondary validation: Sayogo et al. (2014)
d. Guidance	Datasets are accompanied by links or resources that provide user guidance and support	sum datasets with completed CKAN ‘notes’ and CKAN ‘contact’ metadata fields / sum datasets	$\frac{\sum_{i=1}^N (notes_i + contacts_i)}{\sum_{datasets=1}^N datasets}$ <i>notes</i> : Datasets with completed CKAN ‘notes’ metadata field	Primary source: Simperl and Walker (2019) Secondary validation: Olsina (1999 in Bauer and Scharl, 2000); Huang and Benyoucef (2013); Susser and Ariga (2006); Jaiswal et al. (2010); Liu and Arnett (2000); Flavian et al. (2006)
e. Examples	Examples of re-use (fictitious or real) are provided	As there is no standard format for dataset use case examples, this metric could not be translated to computational form		

<sup>1</sup>The web crawler was configured to identify social share code from Facebook (2019) and Twitter (2019).

<sup>2</sup>The web crawler was configured for each portal to run to the dataset level page depth (in most cases three sub-folders deep).

<sup>3</sup>CKAN typically provides three contact fields: ‘contact name’, ‘contact phone’, ‘contact email’.

<sup>4</sup>*N* denotes population level calculation (i.e. *N* = all datasets); *n* denotes sample or subset calculation (i.e. *n* = sample or subset of webpages).

Table 8: Proposal of formulas to calculate the metrics around being accessible, from Dix (2019)..

Be accessible: “data is available to the widest range of users for the widest range of purposes” (Simperl and Walker, 2019, p20):

- The maximum possible score for metric 10b is ‘1’ i.e. 100% of datasets;
- The maximum possible score for metric 10c is also ‘1’ i.e. ‘>1 lang attribute value=1’ or ‘1 lang attribute value=0’;
- The maximum possible score for metric 10e is also ‘1’ i.e. 100% averaged across the four composite metrics;
- The maximum mean score for this category is therefore also ‘1’.

Metric (Simperl and Walker, 2019)	Natural language description (Simperl and Walker, 2019)	Calculation <sup>1,2</sup>	Notation <sup>3</sup>	Source
a. Human and machine readable non-proprietary formats	The portal uses human and machine-readable and non-proprietary formats (e.g. CSV, XML, RDF based formats)	This metric is addressed by category 7		
b. Different formats for the same datasets	The portal provides different types of formats for the same dataset	sum datasets available in >1 file format / sum datasets	$\frac{\sum_{multiformat=1}^N multiformat}{\sum_{datasets=1}^N datasets}$ <i>multiformat</i> : Datasets available in more than one file format <i>datasets</i> : Total portal datasets	Primary source: Simperl and Walker (2019) Secondary validation: NA
c. Instructions provided	The mechanisms for accessing and interacting with datasets are documented	This metric is addressed by metric 2d		
d. Multilingual support	Multilingual support is available on the portal	portal has >1 lang attribute value (1/0)	$n_{lang>1} \Rightarrow multilingual = 1$ $n_{lang=1} \Rightarrow multilingual = 0$ <i>lang</i> : lang attribute value <i>multilingual</i> : Multilingual support metric score	Primary source: Simperl and Walker (2019) Secondary validation: Olsina (1999 in Bauer and Scharl, 2000); Sayogo et al. (2014); W3C (2016)
e. Visual and hearing impaired support	The portal supports the visually and hearing impaired	1. sum images with <img> alt attribute text / sum images	$a = \frac{\sum_{alt=1}^n alt}{\sum_{images=1}^n images}$ <i>alt</i> : Images with <img> alt attribute text <i>images</i> : Total images	Primary source: Simperl and Walker (2019) Secondary validation: Olsina (1999 in Bauer and Scharl, 2000); Ability Net (2018); Hasan and Abuelrub (2011); Mozilla (2019); W3C (2019; 2015)
		2. sum webpages with <title> text / sum webpages	$b = \frac{\sum_{title=1}^n title}{\sum_{pages=1}^n pages}$ <i>title</i> : Webpages with <title> text <i>pages</i> : Webpages crawled	
		3. sum webpages with <h1> and <h2> text / sum webpages	$c = \frac{\sum_{head=1}^n head}{\sum_{pages=1}^n pages}$ <i>head</i> : Webpages with <h1> and <h2> text	
		4. sum webpages with >=1 HTML5 semantic tags / sum webpages	$d = \frac{\sum_{HTML=1}^n HTML}{\sum_{pages=1}^n pages}$ <i>HTML</i> : Webpages with >=1 HTML5 semantic tags	
		mean average score for points 1,2,3,4	$\frac{(a + b + c + d)}{4}$	

<sup>1</sup> The web crawler was configured for each portal to run to the dataset level page depth (in most cases three sub-folders deep).

<sup>2</sup> HTML5 semantic tags included in the web crawl included: <article>|<aside>|<details>|<figcaption>|<figure>|<footer>|<header>|<main>|<mark>|<nav>|<section>|<summary>|<time>.

<sup>3</sup> N denotes population level calculation (i.e. N = all datasets); n denotes sample or subset calculation (i.e. n = sample or subset of webpages).

Dix (2019) evaluates 12 open data CKAN portals based on this measurement methodology and issues a number of recommendations, for some of the categories/dimensions identified in Publications Office of the European Union (2020).

- **Organise for use.** A small proportion of datasets provide web-based dataset previews; portals using CKAN as a deployment platform can enable this functionality by making use of the ‘Datastore’ extension.
- **Promote use.** Contact details are missing, on average, for around one quarter of datasets; developing an appropriate governance framework to ensure that at least one mode of contact – email or phone – is provided to every dataset upload will help to close this gap.
- **Promote use.** Social share code is missing, on average, from around half of dataset or equivalent subfolder-level webpages; deployment at CKAN or CMS template level, or migration to a tag management platform, will allow comprehensive deployment of third-party technology scripts.
- **Be discoverable.** A large proportion of datasets with scheduled update frequencies were found to be out of date; a full audit of existing datasets followed by amendments to metadata or implementation of required updates is necessary.
- **Promote standards.** For all portals, explicit versioning information is absent from the majority of datasets or entirely missing; portal-level implementation and governance strategies, aligned with DCAT-AP guidelines, are required to address this.
- **Co-locate documentation.** For all portals, explicit information on provenance is absent from the majority of datasets or entirely; portal-level implementation and governance strategies are required to address this.
- **Linked data.** Provision of data to RDF standards is still the exception rather than the norm; if open data is to realise increased use value, continued effort to migrate towards a linked-data model is required.
- **Be measurable.** Web analytics tracking is missing for some webpages for a minority of portals; as per the solution for social share code deployment, deployment at CKAN or CMS template level, or migration to a tag management platform will allow the comprehensive deployment of third-party technology scripts.
- **Be accessible.** While findings for ‘be accessible’ are inconclusive, absence of multilingual support and HTML5 semantic tags suggests a lack of adherence to basic accessibility norms by many portals. Accessibility has not been comprehensively addressed here, but a focused effort on behalf of researchers and portal providers in this area is required if portals are to conform to the web content accessibility guidelines of the World Wide Web Consortium.

Publications Office of the European Union (2020) proposes a methodology focused on the ‘be measurable’ dimension of Table 9. More specifically, a proposal was made for open data portals to work on a methodology that would allow to capture reuse indicators. The methodology started with the assessment scoping and definition of reuse metrics in line with the capabilities of the portal and underlying technical infrastructure, followed by the collection of reuse metrics or proxies at various levels of aggregation and the definition of actionable reuse indicators. A list of indicators was already provided, based on a comprehensive survey of literature of data reuse guidance, going beyond open government data. This list is provided in Table 9, inspired by some of the observable features from other platforms where data is also used, such as GitHub. Not all features are directly related to data

reuse. Some features, such as the number of missing values, give insight into the quality of the dataset, instead of its usage.

Table 9: Impact assessment indicators that may be obtained via automated means (Publications Office of the European Union, 2020).

Category of feature	Feature
<b>Portal</b>	Size of repository
	Number of all data files
	Licence
	Dominant data filetype (number of CSV and other files)
	Description
	Ratio of open to closed issues
	Ratio of data files to all files in a repository
	Problematic files with respect to a particular library
<b>Documentation/metadata</b>	Length of the documentation
	Unique URLs
	Language of the documentation
	Number of coding blocks (i.e. both inline and highlighting blocks)
	Number of images
	Broken URIS
<b>Data files</b>	Number of rows and columns of each individual data file
	Missing values
	Data type of headers (i.e. check if headers are strings)
	Size of each data file
	Aggregated size of all the data files in the repository

### 4.3 A systematic literature review on automation of impact indicators

The content of the previous section was based on previous research carried out as part of the extensive study on the sustainability of open data portals and on data.europa.eu research on initial attempts to automate metrics related to the certain specific dimensions it identified.



This section performs a more systematic study of existing work, with the aim of identifying references in the literature where the automation of impact indicators has been discussed. This systematic literature review considers Web of Science as the main source for documents in the scientific domain. The goal of this review is to identify relevant papers discussing computational or manual techniques that may be used to calculate impact indicators.

## Selection of sources

The literature search was carried out by using the following query on topics or titles: ('impact indicator' OR 'metric') AND 'open data'. The title, abstract and keywords (both author-generated and automatically generated) are considered to establish the relevance of the source. For reproducibility purposes, the query used on Web of Science is provided as a footnote <sup>(134)</sup>.

This search resulted in 264 Web of Science papers. Unsurprisingly, a large number of papers identified with this query overlap with those identified in Section 1, although Section 4 does not restrict the search to the last 3 years. First, the title and abstract were reviewed, using the following inclusion criteria: (1) the document should contain clear references to the impact indicators or metrics around open government data; and (2) the document should discuss how the indicators can be calculated. The second condition was important to filter in those documents that were actually relevant for this study, since we are interested in how the indicators are calculated (and later how the calculation can be automated for some of them). Where title and abstract were inconclusive, the introduction and conclusion sections were considered.

From the initial set of 264 papers, 10 papers were retained for a more detailed reading, since they seemed to clearly address impact indicators and how to calculate them, automatically or not (these are marked as 'YES', for 3 papers), or although it was not completely clear, they seemed to propose some types of impact indicators or some ways to calculate them, and required further reading (7 papers, marked as 'MAYBE'). It is important to note that from the initial list several papers that had been identified with the proposed query in Web of Science were discarded since they dealt with metrics that were not focused on impact but mostly on the quality of datasets (on the supply side), for which many papers have been published. There were also several papers discussing different types of techniques associated with linked data publishing, but not necessarily with open data portals, since the term 'linked open data' has been traditionally used quite widely.

The 10 papers that were read completely are the following:

- Johnson (2016),
- Quarati and De Martino (2019),
- Silva and Kerr Pinheiro (2018),
- Abella, Ortiz de Urbina Criado and De Pablos Heredero (2018),
- Susha et al. (2015),
- Raça et al. (2021),
- Planas and Baneres (2018),
- Zeng and Clunis (2020),
- Robinson-Garcia et al. (2017),
- Stuart (2017).

---

<sup>(134)</sup> <https://www.webofscience.com/wos/alldb/summary/7edba31b-51a8-45fd-8cc5-0112282d530d-4289c48b/relevance/1>

The first seven publications are focused on open government data, while the last three were selected because of their potential to provide alternative impact indicators from different areas (linked data and data citation and bibliometrics).

Finally, the references of these papers were considered in order to identify additional papers that were not on the list above (by using snowballing). Moreover, very recent work that had not been published yet is taken into consideration, such as the upcoming publication from Herrera-Murillo et al. (2022).

### Analysis of selected papers and proposal of indicators that can be automated

The goal of the analysis is to obtain a catalogue of the indicators and metrics discussed in the papers and understand how these could be related to open government data initiatives. Some of these papers' findings coincide with the analysis presented in Section 4.2, hence reinforcing some of the initial conclusions. Also, some papers did not provide sufficient details on new indicators or on the techniques used for automating their calculation, and therefore are not discussed much further (a brief explanation of their content is provided at the end of this section).

The findings are grouped in the following four groups: basic metrics around portal usage and interaction with datasets, metrics for open data APIs, metrics associated to the use of open government data in scientific contexts, and finally metrics associated with the mentions of datasets in social networks and other external sources.

#### **Basic metrics around the usage of the portal and interaction with datasets**

Several of the papers that have been analysed discuss how to calculate metrics around the usage of the portal and the interaction with datasets. Johnson (2016) discusses the findings resulting from the interviews of several people responsible of municipal open data portals in Canada, and highlights the so-called passive forms of tracking usage that consist in dataset downloads and other metrics alike, although the paper does not provide sufficient detail on how such evaluation can be automated. They also consider active forms of tracking that are not automated, such as engaging with data reusers more actively.

Raça et al (2021) presents an implemented framework that is able to interact with CKAN-based open data portals in order to calculate some basic metrics. More specifically, the metrics that are calculated are also very simple: number of datasets, number of organisations, number of groups, number of licenses, dataset file format types, publishers' names, public sector bodies and types of licenses. As we can see, these are metrics that really focus on a certain aspect of data quality on the provider side, and do not assess the potential usage of datasets (e.g. downloads), so this is not so relevant for our study.

Abella et al. (2018) analyse several indicators from open data portals, including aspects related to how the datasets follow the 5-star Linked Data model, the number of downloads and the thematic topic of each dataset according to the Spanish classification for open datasets. From these indicators, the only relevant one is the number of downloads, which we have already discussed. No further proposals of automatable metrics are provided.

Zeng and Clunis (2020) focus the discussion not specifically on open government data, but on a very specific type of data resource- SKOS vocabularies. This paper was selected because it mentioned in its abstract that it defines several metrics for determining impact for this specific type of resource (SKOS vocabularies), and that this can be done automatically. These metrics are the following: exposure

through terminology services (facilitating its usage and findability), usage by data providers (based on the inclusion in the metadata of datasets), mapping with other vocabularies (facilitating interoperability) and discussion at professional conferences and in publications (which is related to what we will discuss later on the usage of datasets in science). While not all of these indicators are strictly related to calculating impact, they are all automatable and can provide interesting insights on the usage and impact of these resources, which may be extrapolated to datasets.

Finally, Quarati and de Martino (2019) analyse and make recommendations on the metadata about dataset usage that may be included in open government data portals. Their work relies on data providers to provide this type of information, so that it can be dealt with automatically. They carry out a specific analysis of the number of online views and the number of downloads associated to every portal dataset. Indeed, they explain that these metrics only allow the behaviour of direct users to be determined, while ‘a more mature measure to assess the impact of datasets on end users could take into consideration the indirect users, those who use data indirectly, i.e. processed by intermediaries. These values cannot be inferred from the current portals.’

### Metrics for open data APIs

Not all datasets in open data portals are provisioned by means of APIs, even though the recommendations from the open data directive for high-value datasets will probably have an impact on the availability of APIs. Therefore, not many publications discuss this in the context of open government data. We have selected the work presented in Planas and Baneres (2018), which can be extrapolated to open government data. It proposes two types of metrics that can be applied to the case of open-data APIs (and hence are not fully applicable to all open data portals, where in many cases APIs are not available): performance metrics and semantic metrics. Performance metrics are those that can be calculated directly at the API level:

- average response time by API;
- average response time by request/sub-request;
- average number of accessed APIs and generated sub-requests for each request;
- API reliability;
- query history.

These metrics may not be too representative of impact, but they can provide some insight into how data is being used, especially the query history case.

The semantic metrics may be more relevant. They show the ratio of consumed data within the global model: based on the model that represents the relationship among data stored in one API or several APIs, semantic metrics help to analyse which entities and fields are most in demand. Among these, the following metrics are proposed.

- **Heat UML model.** This metric transforms the model into a heat map where entities’ consumption is highlighted using different colors. This metric will be useful to evaluate which data is mostly accessed.
- **Entity/field consumption.** Additionally, the number of requests for a specific entity or field is shown quantitatively in the model by general requests and sub-requests. This metric is a fine-grained view of previous metrics showing which fields are mainly used.

- **Query diagram.** The open data protocol allows to merge information from different entities in a single request. This metric focuses on showing the navigability to resolve the performed query.

### Use of open government data in (open) science

In the context of papers focusing on data citation in open science, two selected papers are relevant and may be helpful for public administrations to automatically determine how their open datasets are being used in the context of scientific publications.

In Robinson-García et al. (2017), the possibilities of DataCite as a source for bibliometrics is discussed. The authors point out that ‘for data citations to become a valid indicator on data reuse, a shift is needed on the communication behaviour of researchers when citing sources, as well as on the meaning they attach to their references’. If we extrapolate this to the use of open government data sources, two main changes are required. First, open government data publishers should provide a means to associate persistent identifiers (e.g. digital object identifiers) to the datasets that they publish, as discussed in recommendation 10 of Publications Office of the European Union (2022b), so as to facilitate the tracking of their mentions in the literature. Such a tool is already in place for some datasets in data.europa.eu, as discussed in the aforementioned document, as well as by some municipalities who have started applying these recommendations (e.g. the city of Zaragoza in their community site at Zenodo <sup>(135)</sup>, although other systems such as DataCite may be used for this purpose). Second, researchers making use of open government datasets should be made aware of the existence of these persistent identifiers and use them formally, in the reference section, instead of the URLs of the datasets as footnotes or inside the text. This way, the usage of such datasets in scientific publications may be calculated automatically using the means provided by these services (e.g. the DataCite REST API <sup>(136)</sup>, the Zenodo REST API <sup>(137)</sup>, the OpenAIRE Research Graph <sup>(138)</sup>). An important aspect to note is that most of the metadata items required by these systems (Zenodo, DataCite, etc.) are compliant with those filled-in in DCAT-AP. This could be similar for other sources such as Google Scholar, as discussed by Herrera-Murillo et al. (2022), although the authors do the analysis manually in their paper.

### Mentions of open government data in social networks

In the recent paper of Herrera-Murillo et al (2022), resulting from the ODECO doctoral network <sup>(139)</sup>, the authors explore the activity generated by national open data initiatives in Europe in a social network such as Twitter during the year 2021, by obtaining data from the Twitter API. Some of the variables considered in this study and for which automation was provided were the number of relevant tweets, the number of tweets by portal account, the number of users posting tweets and the number of interactions generated by tweets (sum or retweets, replies, quotes and likes). They also explore the mentions of open government datasets in the scientific space in Google Scholar, with the metric of number of items in Google Scholar citing the portal, although these values are obtained manually. Then

---

<sup>(135)</sup> <https://zenodo.org/communities/zgz/>

<sup>(136)</sup> <https://api.datacite.org/>

<sup>(137)</sup> <https://developers.zenodo.org/>

<sup>(138)</sup> <https://graph.openaire.eu/>

<sup>(139)</sup> <https://odeco-research.eu/>

all these values are used to provide a characterisation of different portals, finding 5 main clusters among the 27 national open data portals explored.

#### **Publications discarded after full-text reading**

As discussed above, several publications were discarded after full-text reading, since they did not provide any detailed discussion on how to automate the calculation of metrics or indicators.

- Silva and Kerr Pinheiro (2018) review existing frameworks for open government data but do not provide specific indicators associated with impact assessment, despite what was discussed in the abstract.
- Susha et al. (2015) discuss five open data benchmarks available at the time of writing the publication: the Open Data Readiness Assessment, the Open Data Barometer, the Open Data Index, the PSI Scoreboard and the Open Data Economy benchmarking research (by Capgemini Consulting). It focuses on analysing their methodologies and putting them in the context of academic theories, but does not provide any discussion on the automation of metrics.

## 4.4 Conclusion: considerations for the automation of impact indicators

The literature review, as discussed in Sections 4.2 and 4.3, shows the state of play regarding impact assessment of open data, with a focus on which indicators may be obtained automatically and therefore allow for a more efficient and up-to-date calculation. The reviewed papers show that the number of dataset downloads can be generated automatically relatively easily. Regarding API metrics, the limited relevant literature provides examples such as a Heat UML Model, a heat map which allows to evaluate which data is mostly accessed. Finally, the literature shows how to gain insight on where a dataset is used in science or mentioned on social media.

Based on the literature review, some recommendations are provided for data providers and intermediaries in order to facilitate the process of automatically obtaining impact indicators.

In general, two groups of approaches can be distinguished, depending on whether this is done by data publishers or by intermediaries.

- On the supply side, publishers should continue providing explicitly, as part of the metadata of their datasets and catalogues, basic metadata related to the number of views and downloads associated with each dataset. They should also provide, if possible, as much data as possible obtained from website analytics methods, which generate logs that can be useful to understand user behaviour, and inform process and platform improvements. Finally, they may also decide to provide APIs to access their data, together with the metrics associated with their usage, since APIs provide many more possibilities for reporting data usage, as explained by Planas and Baneres (2018).
- On the demand side, data intermediaries may also develop crawlers, harvesters, and search engines that provide appropriate data identifiers and metadata. These crawlers and harvesters may check data usage in scientific literature and in social networks, among other things, as described in some of the papers that have been discussed. Search engines themselves could

use analytics to gain insight into queries, search patterns, user journeys, and improve dataset use.

More specifically, we recommend that data providers and intermediaries add (using automated tools) the following elements to their open data platforms, as part of the explicit metadata provided with their datasets.

- Number of downloads and visits for each dataset.
- Aggregated number of downloads and visits for each dataset on the different portals where it is published (e.g. those from a municipality that are aggregated by their corresponding national data portal and finally by data.europa.eu).
- Availability of datasets in other general-purpose or community-specific services. For instance, availability of the datasets in GitHub, Zenodo, thematic data portals (e.g. environment), search engines, etc.
- Availability of libraries (e.g. in GitHub repositories, specific Python libraries) that allow making use of the datasets.
- Explicit links among datasets (outgoing and incoming) and with existing vocabularies.
- Generation of persistent identifiers and times that a dataset is referenced in scientific literature, in social networks or in other information resources using those identifiers.
- Query history for dataset search in open data portals.
- Query history on API calls in the data APIs of open data portals or services.
- Statistics of consumption of API calls, with details on requested resources and requested attributes/fields.

Many of the indicators proposed in Section 2 and Section 3 can thus be measured with automated tooling if open data portals and open data intermediaries enable and publish certain (meta) data. Automation could play a valuable role in measuring open data impact after a clear connection between the number of use cases (or users) and actual impact has been established.

## 5 Conclusions

This report is the first in a series of four that should establish a common method to assess the impact of open data. As a first step, this report provides an overview of existing academic literature on the impact of open data, the impact of open data on open data portals, the impact via intermediaries and ways to measure impact with automated tooling.

Assessing the impact of open data is particularly challenging given its breadth. By definition, it potentially reaches every corner of the European economy and society. Impact assessments often have a more narrowly defined impact area. For instance, an impact assessment report estimated that the EU ecodesign regulation had saved some 1 037 terawatt hours of primary energy by 2020 <sup>(140)</sup>. Although ecodesign impact assessment is an enormous task, the methodology for the impact assessment itself is straightforward: the goal of the regulation is to save energy and the impact assessment evaluates how much energy is saved.

Additionally, measuring the impact of open data is challenging because it is hard to pinpoint what it is that needs to be evaluated. Instead of assessing the impact of the open data directive, the challenge is to evaluate all impact that is created with open data. Consequently, the impact of open data cannot be attributed to a specific sector or domain. Since open data can be about anything, its impact can too. How do you compare a transportation app that saves time and a medicine app that improves public health with economic benefits?

The literature reflects this ubiquity as well, as contemporary frameworks to measure the impact of open data currently used by countries, academia and other stakeholders use different definitions for the impact of open data that often evaluate social, political, economic and environmental impacts. What is needed is a common definition of open data that is able to reflect the multitude of ways in which open data achieves impact. In the absence of a common definition, we recommend structuring the impact assessment in line recommendations on how they can be structured, as described in the better regulation guidelines and toolbox <sup>(141)</sup>, as they serve as the golden standard in the EU when preparing new initiatives. This means that the impact of open data would be classified as social, economic and environmental impact.

The lack of a common framework does not mean that the impact of open data is currently not being measured. Quantitative data about the impact of open data is available. User statistics, such as the number of users, the number of downloads and the number of reuse cases on open data portals and open data intermediaries provide bigger and smaller breadcrumbs to measure the path from open data to impact. Moreover, these breadcrumbs can often be tracked in automated ways. The downside is that user statistics on open data portals and open data intermediaries do not actually measure impact. User statistics are valuable proxies that provide an indication of the impact of data, but the precise way in which impact is generated remains unclear.

There exists a wealth of qualitative information on open data use cases that can complement the quantitative user statistics. The main challenge ahead is to find ways to connect these data points to

---

<sup>(140)</sup> Ecodesign impact accounting annual report 2020: <https://op.europa.eu/en/publication-detail/-/publication/568cac02-5191-11ec-91ac-01aa75ed71a1/language-en>.

<sup>(141)</sup> [https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how/better-regulation-guidelines-and-toolbox\\_en](https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how/better-regulation-guidelines-and-toolbox_en)

obtain genuine insights about the impact of open data. The question is not just how often a dataset has been used for reuse, but also what the impact of the reuse case was on society. Open data research frameworks such as the ODM recognise this need too: instead of merely providing reuse cases, countries are asked to explain how impact is created through these use cases.

If the impact of specific use cases could be established in a consistent matter, then the combination of user statistics and use cases would be a solid fundament for an open data impact assessment. However, (financial) data of use cases needs to be evaluated to assess the impact of open data use cases. As it stands, many (commercial) open data users keep such KPIs to themselves, since it might serve as a competitive advantage to others.

Further research into the representativeness and completeness of use cases is required to be able to go from estimating the impact of a specific use case to knowing what the general impact of open data is. If information about the number of use cases is available, along with impact estimates for a number of use cases, then the general impact of open data could be extrapolated from these estimates.



## Bibliography

1. Abella, A., Ortiz de Urbina Criado, M. and De Pablos Heredero, C. (2018), 'Indicadores de calidad de datos abiertos: el caso del portal de datos abiertos de Barcelona', *Profesional de la información*, Vol. 27, No 2, pp. 375–382 (<https://doi.org/10.3145/epi.2018.mar.16>).
2. Ackerman, D. (2021), 'When more Covid-19 data doesn't equal more understanding,' *MIT News*, Massachusetts Institute of Technology (<https://news.mit.edu/2021/when-more-covid-data-doesnt-equal-more-understanding-0304>).
3. ASEDIE (2021), '2021 infomediary sector' (<https://www.asedie.es/en/annual-report>).
4. ASEDIE (2022), 'Economía del dato en el ámbito infomediario' (<https://www.asedie.es/es/informes-anuales>).
5. Apanasevic, T. (2021), *Socio-economic effects and the value of open data: A case from Sweden*, International Telecommunications Society (ITS), Calgary (<https://www.econstor.eu/handle/10419/238004>).
6. Belickas, J. (2022), '2021 metų viešojo sektoriaus duomenų atvėrimo pažangos apžvalga' (<https://data.gov.lt/news/23>).
7. Berends, J., Carrara, W. and Radu, C. (2020), *Analytical Report 9: The economic benefits of open data*, Publications Office of the European Union, Luxembourg ([https://data.europa.eu/sites/default/files/analytical\\_report\\_n9\\_economic\\_benefits\\_of\\_open\\_data.pdf](https://data.europa.eu/sites/default/files/analytical_report_n9_economic_benefits_of_open_data.pdf)).
8. Benedetti, M., Russo, C., Tangi, L. and Vanini, I. (2020), 'Where does the value of open data lie? A framework to assess open data impacts'.
9. Bhattacharya, J., Bhattacharyya, R., Singla, H. and Vigneswara Ilavarasan, P. (2022), 'Estimation of economic impact of national open digital ecosystems', in Senjyu, T., Mahalle, P., Perumal, T. and Joshi, A. (eds), *ICT with Intelligent Applications*, Springer, pp. 25–32 ([https://doi.org/10.1007/978-981-16-4177-0\\_4](https://doi.org/10.1007/978-981-16-4177-0_4)).
10. Van Belle, J.-P. and Hlabano, M. (2019), 'Building urban resilience in the face of severe drought through the innovative use of open data', *IEEE*, pp. 1–7 (<https://doi.org/10.1109/ICBDSC.2019.8645586>).
11. Cannon, M., Kelly, A. and Freeman, C. (2022), 'Implementing an open & fair data sharing policy – a case study in the earth and environmental sciences', *Learned Publishing*, Vol. 35, No 1, pp. 56–66 (<https://doi.org/10.1002/leap.1442>).
12. Centre for Data Ethics and Innovation (2021), 'Unlocking the value of data: exploring the role of data intermediaries' ([https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1004925/Data\\_intermediaries\\_-\\_accessible\\_version.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1004925/Data_intermediaries_-_accessible_version.pdf)).
13. Chapman, A., Drake, A., Huynh, D., Hyde, R., Keller, P., Kleinman, M., Lisinska, J., Moreau, L., Mousavi, M. R., Naiseh, m., Ramchurn, S., Sempredoni, D., Simperl, E., Stalla-Bourdillon, S., Thuermer, G., Tsakalakis, N., Tyukin, I. and Weerawardhana, S. (2021), 'Data: a new direction – a call for evidence from department for digital, culture, media and sport', King's College London (<https://doi.org/10.18742/pub01-069>).
14. Chu, P.-Y. and Lee, F.-W. (2019), 'An evaluation of motivations and perceived impacts of open government data', 17th International Conference e-Society, pp. 167–174 ([https://doi.org/10.33965/es2019\\_201904L021](https://doi.org/10.33965/es2019_201904L021)).
15. Corrales-Garay, D., Ortiz de Urbina Criado, M. and Mora-Valentín, E.-M. (2020), 'A research agenda on open data impact process for open innovation', *IEEE Access*, Vol. 8 (<https://doi.org/10.1109/ACCESS.2020.2974378>).
16. Crusoe, J. and Melin, U. (2018), 'Investigating open government data barriers,' in Parycek, P., Glassey, O., Janssen, M., Scholl, H. J., Tambouris, E., Kalampokis, E. and Virkar, S. (eds), *Electronic Government*, Springer International Publishing, pp. 169–183 ([https://doi.org/10.1007/978-3-319-98690-6\\_15](https://doi.org/10.1007/978-3-319-98690-6_15)).

17. Davies, T. Walker, S., Rubinstein, M. and Perini, F. (2019), 'The state of open data: histories and horizons', African Minds and the International Development Research Centre, Cape Town and Ottawa.
18. Deloitte (2021a), 'Conducting an open data impact study in Cyprus and developing a measurement methodology for the open data impact and analysis report in Cyprus', Deloitte Limited, Nicosia.
19. Deloitte (2021b), 'Conducting an open data impact study in Cyprus and developing a measurement methodology for the handbook "Good Practices and Methodology for Evaluating the Use of Open Data"', Deloitte Limited, Nicosia.
20. den Haan, D. (2018), 'Intermediation as a dissolver of barriers: how intermediaries can help overcome barriers in open data use', *Utrecht University* (<https://studenttheses.uu.nl/handle/20.500.12932/31148>).
21. Dix, M. (2019), 'Measuring open data portal user-orientation: a computational approach', University of Southampton.
22. European Commission (2019), Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on Open Data and the Re-Use of Public Sector Information, Official Journal of the European Union (<https://eur-lex.europa.eu/eli/dir/2019/1024/oj>).
23. European Commission (2022), Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), *Official Journal of the European Union* (<http://data.europa.eu/eli/reg/2022/868/oj>).
24. G8 (2016), 'G8 Open Data Charter', *International Open Data Charter* (<https://opendatacharter.net/g8-open-data-charter/>).
25. Gao, Y. and Janssen, M. (2022), 'The open data canvas – analyzing value creation from open data', *Digital Government: Research and Practice*, Vol. 3, No 1, pp. 1–15 (<https://doi.org/10.1145/3511102>).
26. Herrera-Murillo, D. J., Aziz, A., Noguera-Iso, J. and Lopez-Pellicer, F. J. (2022), 'Analysing user involvement in open government data initiatives', in Silvello, G., Corcho, O., Manghi, P., Di Nunzio, G. M., Golub, K., Ferro, N. and Poggi, A. (eds), *Linking Theory and Practice of Digital Libraries*, Springer International Publishing, pp. 175–186.
27. Hlabano, M. and Van Belle, J.-P. (2019), 'Tracing the impact of the city of cape town's open data initiative on communities and development' in *International Conference on Social Implications of Computers in Developing Countries*, Springer, pp. 284–294.
28. Hrustek, L., Tomičić Furjan, M. and Pihir, I. (2021), 'Political participation in the information society and impact of open data on it', IEEE Technology & Engineering Management Conference-Europe (TEMSCON-EUR), pp. 1–6.
29. Hsu, J., Ravichandran, R., Zhang, E. and Keung, C. (2021), 'Open data standard and analysis framework: towards response equity in local governments', *Equity and Access in Algorithms, Mechanisms, and Optimization*, pp. 1–8.
30. James, P. (2022), 'Civic technology: a chutes and ladders analysis of transformative potential', *Journal of Smart Cities and Society*, Vol. 1, No 1, pp. 3–15 (<https://doi.org/10.3233/SCS-210117>).
31. Janssen, M. and Zuiderwijk, A. (2014), 'Infomediary business models for connecting open data providers and users', *Social Science Computer Review*, Vol. 32, No 5, pp. 694–711 (<https://doi.org/10.1177/0894439314525902>).
32. Johnson, P. and Greene, S. (2017), 'Who are government open data infomediaries? A preliminary scan and classification of open data users and products', *URISA Journal*, Vol. 28, No 14, p. 9.
33. Johnson, P. (2016), 'Reflecting on the success of open data: how municipal government evaluates their open data programs', *International Journal of E-Planning Research*, Vol. 5, pp. 1–12.

34. Kawashita, I., Baptista, A. A. and Soares, D. (2020), 'An assessment of open government data benchmark instruments'.
35. Krishnamurthy, R. and Awazu, Y. (2016), 'Liberating data for public value: the case of Data.Gov,' *International Journal of Information Management*, Vol. 36, No 4, pp. 668–672 (<https://doi.org/10.1016/j.ijinfomgt.2016.03.002>).
36. Lassinantti, J. (2019), 'Reuse of public sector open data – characterising the phenomena', *International Journal of Enterprise Information Systems*, Vol. 13, pp. 1–29.
38. Lateral Economics and ODI (2021), 'Research: The economic value of open versus paid data' (<https://theodi.org/article/research-the-economic-value-of-open-versus-paid-data/>).
39. Lokers, R., Berdou, E., Miguel Ayala, L. and Mey, L. (2019), 'Guidelines for analysing pathways to impact: evaluation of open data for development', *F1000Research*, Vol. 8, No 2115, (<https://doi.org/10.7490/f1000research.1117778.1>).
40. Lokers, R., Miguel Ayala, L. and Berdou, E. (2019), 'Knowledge reservoir for impact evaluation – a collection of resources on impact evaluation of open data for development', *F1000Research*, Vol. 8, No 2121.
41. Loshin, D. (2013), 'Chapter 18 – Repurposing publicly available data', in Loshin, D. and Kaufmann, M. (eds), *Business Intelligence*, pp. 287–301 (<https://doi.org/10.1016/B978-0-12-385889-4.00018-1>).
42. McLean, S., Perez Maestro, L., Hernandez, S., Mezzofiore, G. and Polglase, K. (2021), 'The Covid-19 pandemic has catapulted one mysterious data website to prominence, sowing confusion in international rankings', *CNN* (<https://www.cnn.com/interactive/2020/05/world/worldometer-coronavirus-mystery/>).
43. Ministry of Digital Transformation of Ukraine (2021), 'Anti-corruption and social impact of open data in the field of financial transparency of Ukraine: analytical report', 43 (<https://tapas.org.ua/wp-content/uploads/2021/07/Open-Data-Impact-Report-Financial-Transparency.pdf>) (in Ukrainian).
44. Ministry of Digital Transformation of Ukraine (2021), 'Anti-corruption and social impact of open data in the field of forestry: analytical report', 38 (<https://tapas.org.ua/wp-content/uploads/2021/08/Open-Data-Impact-Report-Forestry.pdf>) (in Ukrainian).
46. Ministry of Digital Transformation of Ukraine (2021), 'Anti-corruption and social impact of open data in the field of state supervision (control): analytical report', 41 (<https://tapas.org.ua/wp-content/uploads/2021/06/Open-Data-Impact-Report-State-Supervision-and-Control.pdf>) (in Ukrainian).
47. Ministry of Economic Affairs and Communications of Ukraine (2021), 'The second impact assessment of open data in the public sector' (<https://avaandmed.eesti.ee/instructions/teistkordne-avaandmete-mojuhinnang-avalikus-sektoris>) (in Estonian).
48. Myeong, S., Ahn, M. J., Kim, Y., Chu, S. and Suh, W. (2021), 'Government data performance: the roles of technology, government capacity, and globalization through the effects of national innovativeness', *Sustainability*, Vol. 13, No 22 (<https://doi.org/10.3390/su132212589>).
49. Nerurkar, A. and Das, I. (2020), 'Open data readiness assessment framework for government projects: indian perspective', *International Conference on Information and Communication Technology for Intelligent Systems*, Springer, Singapore.
50. Neves, F. T., de Castro Neto, M. and Aparicio, M. (2020), 'The impacts of open data initiatives on smart cities: a framework for evaluation and monitoring', *Cities*, Vol. 106 (<https://doi.org/10.1016/j.cities.2020.102860>).

51. OECD (2018), *Open Government Data Report: Enhancing Policy Maturity for Sustainable Impact*, OECD Digital Government Studies, OECD Publishing, Paris (<https://doi.org/10.1787/9789264305847-en>).
52. Osorio-Sanabria, M., Amaya-Fernández, F. and González-Zabala, M. P. (2020), 'Developing a model to readiness assessment of open government data in public institutions in Colombia', proceedings of the 13th International Conference on Theory and Practice of Electronic Governance, pp. 334–340 (<https://doi.org/10.1145/3428502.3428548>).
53. Osorio-Sanabria, M., Brito-Carvajal, J., Astudillo, H., Amaya-Fernández, F. and González-Zabala, M. P. (2020), 'Evaluating Open Government Data Programs: A Systematic mapping study', 2020 Seventh International Conference on eDemocracy & eGovernment (ICEDEG), IEEE, pp. 157–164 (<https://doi.org/10.1109/ICEDEG48599.2020.9096755>).
54. Planas, E. and Baneres, D. (2018), 'Model-Driven Analytics for Open Data APIs', in Pautasso, C., Sánchez-Figueroa, F., Systa, K. and Murillo Rodríguez, J. M. (eds), *Current Trends in Web Engineering*, Springer International Publishing, pp. 176–182 ([https://doi.org/10.1007/978-3-030-03056-8\\_15](https://doi.org/10.1007/978-3-030-03056-8_15)).
55. Park, S. and Gil-García, J. R. (2022), 'Open data innovation: visualizations and process redesign as a way to bridge the transparency-accountability gap', *Government Information Quarterly*, Vol. 39, No 1 (<https://doi.org/10.1016/j.giq.2020.101456>).
56. Prieto, A., Mazón, J. N., Lozano-Tello, A. and Ibáñez, L. D. (2018), 'Supporting open dataset publication decisions based on open source software reuse', CEUR workshop proceedings, Vol. 2062.
57. Priestley, M., Simperl, E., Juc, C. and Anguiano, M. (2022), 'Measuring the impact of publicly funded open innovation programmes: The case of Data Market Services Accelerator', *Open Research Europe*, Vol. 1, No 71.
58. Publications Office of the European Union (2020), *Sustainability of (open) data portal infrastructures: summary overview and key recommendations*, Publications Office of the European Union, Luxembourg (<https://data.europa.eu/doi/10.2830/086159>).
59. Publications Office of the European Union (2021), 'Open data maturity assessment 2021: detailed country scores', *Open Data in Europe 2021*, Publications Office of the European Union, Luxembourg ([https://data.europa.eu/sites/default/files/country\\_scores\\_2021.xlsx](https://data.europa.eu/sites/default/files/country_scores_2021.xlsx)).
60. Publications Office of the European Union (2022a), *Open Data Maturity Report 2021*, Publications Office of the European Union, Luxembourg (<https://data.europa.eu/doi/10.2830/394148>).
61. Publications Office of the European Union (2022b), *Principles and recommendations to make data.europa.eu data more reusable : a strategy-mapping report*, Publications Office of the European Union, Luxembourg (<https://data.europa.eu/doi/10.2830/9342>).
62. Publications Office of the European Union and European Data Portal (2022), *Measuring data demand within the public sector: Discussion paper 2021*, Publications Office of the European Union, Luxembourg (<https://data.europa.eu/doi/10.2830/54555>).
63. Quarati, A. and De Martino, M. (2019), 'Open government data usage: a brief overview', proceedings of the 23rd International Database Applications & Engineering Symposium, Association for Computing Machinery, pp. 1–8 (<https://doi.org/10.1145/3331076.3331115>).
64. Raça, V., Velinov, G., Cico, B. and Kon-Popovska, M. (2021), 'Application-based framework for analysis, monitoring and evaluation of national open data portals', *International Journal of Advanced Computer Science and Applications*, Vol. 12, pp. 26–36 (<https://doi.org/10.14569/IJACSA.2021.0121104>).

65. Reggi, L. and Dawes, S. S. (2022), 'Creating open government data ecosystems: network relations among governments, user communities, NGOs and the media', p. 5.
66. Robinson-Garcia, N., Mongeon, P., Jeng, W. and Costas, R. (2017), 'DataCite as a novel bibliometric source: coverage, strengths and limitations', *Journal of Informetrics*, Vol. 11, No 3, pp. 841–854 (<https://doi.org/10.1016/j.joi.2017.07.003>).
67. Runeson, P., Olsson, T. and Linåker, J. (2021), 'Open data ecosystems – an empirical investigation into an emerging industry collaboration concept', *Journal of Systems and Software*, Vol. 182, p. 111088 (<https://doi.org/10.1016/j.jss.2021.111088>).
68. Saxena, S. (2020), *An Evaluation of the National Open Government Data (OGD) Portal of the United Arab Emirates*, pp. 191–209 ([https://doi.org/10.1007/978-3-030-08277-2\\_12](https://doi.org/10.1007/978-3-030-08277-2_12)).
69. Shaharudin, A. (2020), 'Open government data: principles, benefits and evaluations', Kuala Lumpur: Khazanah Research Institute. License: Creative Commons Attribution CC BY 3.0.
70. Silva, P. N. and Kerr Pinheiro, M. (2018), 'Métricas para dados governamentais abertos', *Revista Ibero-Americana de Ciência da Informação*, Vol. 11, No 1, pp. 314–332 (<https://doi.org/10.26512/rici.v11.n1.2018.8431>).
71. Susha, I., Zuiderwijk, A., Janssen, M. and Grönlund, A. (2015), 'Benchmarks for evaluating the progress of open data adoption: usage, limitations, and lessons learned', *Social Science Computer Review*, Vol. 33, No 5, pp. 613–630 (<https://doi.org/10.1177/0894439314560852>).
72. Stuart, D. (2017), 'Data bibliometrics: metrics before norms', *Online Information Review*, Vol. 41 (<https://doi.org/10.1108/OIR-01-2017-0008>).
73. Trindade, F., Castro Neto, M and Aparicio, M. (2020), 'The impacts of open data initiatives on smart cities: A framework for evaluation and monitoring', *Cities*, Vol. 106, 102860.
74. Van Schalkwyk, F., Canares, M., Chattapadhyay, S. and Andrason, A. (2015), 'Open data intermediaries in developing countries', Figshare (<https://doi.org/10.6084/M9.FIGSHARE.1449222>).
75. Van Schalkwyk, F., Willmers, M. and McNaughton, M. (2016), 'Viscous open data: the roles of intermediaries in an open data ecosystem', *Information Technology for Development*, Vol. 22, No sup1, pp. 68–83.
76. Verhulst, S. and Young, A. (2017), 'Open data in developing economies: toward building an evidence base on what works and how' (<https://papers.ssrn.com/abstract=3141435>).
77. Vogelesang, F. (2022), 'A closer look at data intermediaries and the risk of platformization', *Open Future* (<https://openfuture.eu/blog/a-closer-look-at-data-intermediaries-and-the-risk-of-platformization>).
78. Walker, J., Frank, M. and Thompson, N. (2016), 'User centred methods for measuring the value of open data', *Journal of Community Informatics*, Vol. 12, No 2.
79. Wauters, R. (2014), 'Open and big data in Europe: potential growth drivers, or a massively squandered opportunity?', Tech.Eu (<https://tech.eu/2014/01/29/open-big-data-in-europe>).
80. Wilson, B. and Cong C. (2021), 'Beyond the supply side: use and impact of municipal open data in the U.S.', *Telematics and Informatics*, Vol. 58 (<https://doi.org/10.1016/j.tele.2020.101526>).
81. World Economic Forum (2022), 'Advancing digital agency: the power of data intermediaries' (<https://www.weforum.org/reports/advancing-digital-agency-the-power-of-data-intermediaries/>).
82. World Wide Web Foundation (2018), *Open Data Barometer – Leaders Edition*, Washington D.C. (<https://opendatabarometer.org/doc/leadersEdition/ODB-leadersEdition-Report.pdf>).
83. Zeng, M. L. and Clunis, J. (2020), 'FAIR + FIT: guiding principles and functional metrics for linked open data (LOD) KOS products', *Journal of Data and Information Science*, Vol. 5, No 1, pp. 93–118 (<https://doi.org/10.2478/jdis-2020-0008>).

84. Zheng, L., Kwok, W.-M, Aguaro, V., Qi, X. and Lyu, W. (2020), 'Evaluating global open government data: methods and status', proceedings of the 13th international conference on theory and practice of electronic governance.
85. Zuiderwijk, A., Pirannejad, A. and Susha, I. (2021), 'Comparing open data benchmarks: which metrics and methodologies determine countries' positions in the ranking lists?', *Telematics and Informatics*, Vol. 62 (<https://doi.org/10.1016/j.tele.2021.101634>).
86. Zuiderwijk, A. and Janssen, M. (2014), 'Open data policies, their implementation and impact: a framework for comparison', *Government Information Quarterly*, Vol. 31, No 1, pp. 17–29 (<https://doi.org/10.1016/j.giq.2013.04.003>).

## Appendices

### Annex I – Methodology for inventory of national and local open data portals

#### 1. Selection of national open data portals

As regards the national level, open data portals are investigated according to an aggregated selection method, based on the results of the [Open Data Maturity Index](#) (latest edition: 2021). Starting from the general assumption that top-scoring countries on specific questions have a higher probability of collecting and publishing statistics about the reuse and impact of open data, the following approach is proposed.

First, the aggregate ranking is created based on the following selected questions from the Open Data Maturity Index Table 10.

*Table 10: Selected questions from the 2021 Open Data Maturity Index*

Question
31. At the national level, are there any processes running to estimate the level of reuse of your country's open data (such as monitoring, surveys, web analytics or catalogues of apps that use the data)? (Dimension 2: Open data impact; 2.1 Strategic awareness)
34. Do you have a methodology in place to estimate the impact of open data in your country? (Dimension 2: Open data impact; 2.1 Strategic awareness)
38. Have you or other public bodies launched or performed any activities in the past year to assess the political impact of open data (such as systematic monitoring, commissioning studies, surveys)? (Dimension 2: Open data impact; 2.2 Political impact)
44. Have you or other public sector stakeholders active in this field launched or performed any activities in the past year to monitor the social impact of open data (such as systematic monitoring, commissioning studies, surveys)? (Dimension 2: Open data impact; 2.3 Social impact)
48. Have you or other public sector stakeholders active in this field launched or performed any activities in the past year to monitor the environmental impact of open data (such as systematic monitoring, commissioning studies, surveys)? (Dimension 2: Open data impact; 2.4 Environmental impact)
53. Have you or other public sector stakeholders active in this field launched or performed any activities in the past year to monitor the economic impact of open data (such as systematic monitoring, commissioning studies, surveys)? (Dimension 2: Open data impact; 2.5 Economic impact)
73. Does the national portal have a designated area to showcase use cases? (Dimension 3: Open data portal; 3.1 Portal features)
74. Does the national portal provide the possibility for users to submit their own use cases? (Dimension 3: Open data portal; 3.1 Portal features)

Question
75. Does the national portal reference the data sets that the showcased use cases are based on? (Dimension 3: Open data portal; 3.1 Portal features)
81a. Do you perform any activities to gain insight into the portal's usage (e.g. web analytics, surveys, or analysis of social media feeds) (Dimension 3: Open data portal; 3.2 Portal usage)

The 12 top-scoring countries are selected for further analysis.

In order to validate the underlying assumption of the selection method, the open data portals of the bottom-five countries regarding the presence of a dedicated statistics/impact/reuse section were scanned as well. This resulted in one anomaly: Georgia.

The third source is an open call to the open data community to provide suggestions of relevant data portals to include in the scan. (see Figure 28). The call was launched on 6 June 2022 using the collaboration channel on the data.europa virtual workspace and social media accounts.

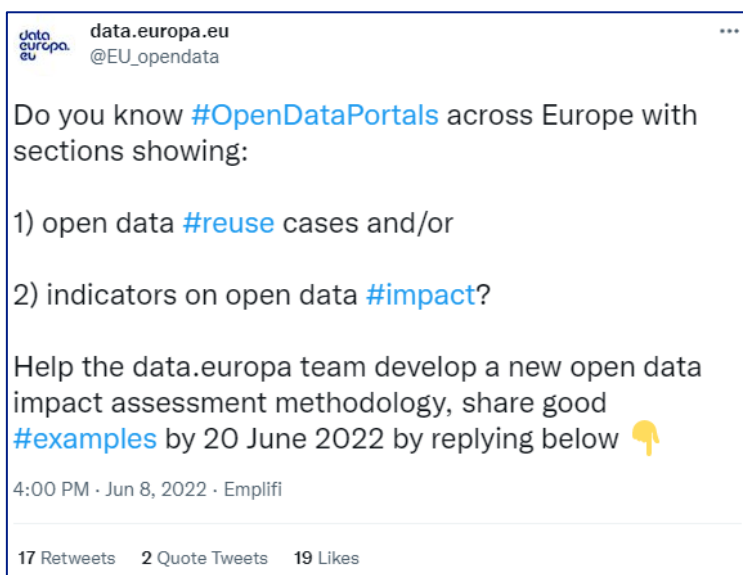


Figure 28: Call for input to the open data community

## 2. Selection of local open data portals

Municipal open data portals were selected in two steps.

- a. A quick scan of 16 key local open data portals selected in previous research reports on open data in cities <sup>(142)</sup>. The quick scan assesses whether the portals contain a section on user statistics, reuse cases or other KPIs related to open data impact. Following this assessment, Barcelona, Berlin, Dublin, Florence, Helsinki region, Lisbon, Paris and Vienna were selected.
- b. Through the community call (see description above), members of the open data community suggested to include the following portals in the analysis: Barcelona, Bordeaux, Eindhoven, Florence, Helsinki region, Madrid and Zaragoza. Barcelona, Florence and the Helsinki region had also been selected in step a.

<sup>(142)</sup> [https://data.europa.eu/sites/default/files/edp\\_analytical\\_report\\_n6\\_-\\_open\\_data\\_in\\_cities\\_2\\_-\\_final-clean.pdf](https://data.europa.eu/sites/default/files/edp_analytical_report_n6_-_open_data_in_cities_2_-_final-clean.pdf);  
[https://data.europa.eu/sites/default/files/edp\\_analytical\\_report\\_n4\\_-\\_open\\_data\\_in\\_cities\\_v1.0\\_final.pdf](https://data.europa.eu/sites/default/files/edp_analytical_report_n4_-_open_data_in_cities_v1.0_final.pdf)



### 3. Analytical framework

The following analytical framework is used for the analysis of the selected portals:

- Reuse case observatory
  - Present (yes/no);
  - Documented characteristics, e.g. type of reuse (app/visualisation/article/report etc); (number of) data sources used; (policy) category; type of reuser (government/business/civil society/data journalist etc.)
  - Employed methodology / instructions to data reusers
  - Place of publication: dedicated page/feature on homepage/ in submenu/featured in dedicated articles, etc.
- Reuse statistics / impact indicators
  - Types, e.g. number of reuses; number of reusers
  - Automated (yes/no)
  - Place of publication: dedicated page/feature on homepage/ in submenu/featured in dedicated articles, etc.
  - Indicators linked to economic impact
  - Indicators linked to social impact
  - Indicators linked to environmental impact
- Reports on open data impact/reuse published on the portal
  - Date
  - Focus
  - Frequency/recurring
  - Methodology/indicators used (including links to economic, social and environmental impact)
- Conclusion: what methods and indicators does a country use to measure and showcase the impact of their open data portal?

### 4. Language considerations

Linguistic barriers must be taken into consideration when analysing open data portals from a large array of European countries. The first tool to take advantage of is the embedded English version of the portals' websites. When these are not available, multilingual neural machine translation services (e.g. Google Translate) are employed.

In addition to this, and to further corroborate the results, the research team boasts an inclusive multicultural team, available to actively help dispel linguistic doubts and provide their expertise as native speakers.

## Annex II – Results of national portals’ quick scans

The results of the quick scan are largely in line with the findings of the ODM aggregate score insofar as a majority of surveyed portals have dedicated use case repositories and use(r) statistics pages. In a small number of cases (e.g. that of Greece), portal administrators indicated on the ODM that processes were running to track portal use and data reuse, but these features could not be found on the portals. It is possible that the results of these analyses are distributed internally but not made available for public viewing. The quick scan of the bottom five countries validated the ODM-based selection method, as little to no impact-focused features were found on these portals.

Country <i>url portal</i>	ODM aggregate score	Reuse case section	Use(r) statistics	Report(s) on impact	Remarks
Cyprus <a href="https://www.data.gov.cy/?language=en">https://www.data.gov.cy/?language=en</a>	200	yes	yes	yes	
Germany <a href="https://www.govdata.de">https://www.govdata.de</a>	200	yes	no	no	The German response to the Open Data Maturity Assessment indicated that the country had implemented mechanisms for tracking portal use, suggesting that these data may be distributed internally but not made available to the public.
Estonia <a href="https://avaandmed.eesti.ee">https://avaandmed.eesti.ee</a>	200	yes	yes	yes	
Spain <a href="https://datos.gob.es/en">https://datos.gob.es/en</a>	200	yes	yes	no	
Finland <a href="https://www.avoindata.fi/en">https://www.avoindata.fi/en</a>	200	yes	yes	no	
France <a href="https://www.data.gouv.fr/en">https://www.data.gouv.fr/en</a>	200	yes	yes	no	
Greece <a href="https://data.gov.gr">https://data.gov.gr</a>	200	no	no	no	The Greek response to the Open Data Maturity Assessment indicated that the country had processes running to track portal use and data reuse, but these

Country <i>url portal</i>	ODM aggregate score	Reuse case section	Use(r) statistics	Report(s) on impact	Remarks
					features could not be found on the portal. It is possible that the results of these analyses are distributed internally but not made available for public viewing.
Croatia <a href="https://data.gov.hr/en">https://data.gov.hr/en</a>	200	yes	yes	no	
Ireland <a href="https://data.gov.ie">https://data.gov.ie</a>	200	yes	yes	no	The Irish Open Data Unit is in the process of conducting an impact study.
Italy <a href="https://dati.gov.it">https://dati.gov.it</a>	200	no	no	no	
Lithuania <a href="https://data.gov.lt/?lang=en">https://data.gov.lt/?lang=en</a>	200	yes	yes	no	The portal has no statistics dashboard, but statistics are compiled in an annual report.
Netherlands <a href="https://data.overheid.nl/en">https://data.overheid.nl/en</a>	200	yes	yes	yes	
Poland <a href="https://dane.gov.pl/en">https://dane.gov.pl/en</a>	200	yes	no	no	The Polish response to the Open Data Maturity Assessment indicated that the country had implemented mechanisms for tracking portal use, suggesting that these data may be distributed internally but not made available to the public.
Ukraine <a href="http://data.gov.ua">http://data.gov.ua</a> <a href="https://diia.data.gov.ua">https://diia.data.gov.ua</a>	200	no	yes	yes	The server on which the portal is hosted does not allow access at the time of writing this report. Some portal features were assessed using an archived version of the statistics page saved on 23 February 2022.
<i>Czechia, Norway, Romania, Sweden, Austria, Bulgaria, Montenegro, Denmark, Luxembourg, Slovenia, Hungary, the UK, Latvia and Portugal have intermediate scores ranging from 120 to 190. Therefore, these countries are not included in the quick scan.</i>					
Belgium <a href="https://data.gov.be/en">https://data.gov.be/en</a>	100	yes	no	no	Visitors to the site may download the results of a 2021 survey of users and contributors in spreadsheet format.

Country <i>url portal</i>	ODM aggregate score	Reuse case section	Use(r) statistics	Report(s) on impact	Remarks
Switzerland <a href="https://opendata.swiss/en">https://opendata.swiss/en</a>	100	yes	no	no	
Malta <a href="https://open.data.gov.mt">https://open.data.gov.mt</a>	70	no	no	no	
Iceland <a href="https://opingogn.is">https://opingogn.is</a>	30	no	no	no	
Georgia <a href="https://www.data.gov.ge">https://www.data.gov.ge</a>	30	yes	yes	no	Although the portal has a designated reuse section, this page has been 'in development' since at least March 2016, and does not list any cases. Portal administrators indicated on the Open Data Maturity Index that they 'did not know' whether processes for estimating reuse were running.
Slovakia <a href="https://data.gov.sk/en">https://data.gov.sk/en</a>	20	yes	no	no	

## Annex III – Results quick scan local portals

City/Region	Reuse case section	Use(r) statistics	Report(s) on impact	Remarks
Amsterdam <a href="https://data.amsterdam.nl">https://data.amsterdam.nl</a>	no	no	no	
Barcelona <a href="https://opendata-ajuntament.barcelona.cat/en">https://opendata-ajuntament.barcelona.cat/en</a>	yes	yes	no	Also proposed through the call to the open data community.
Berlin <a href="https://daten.berlin.de">https://daten.berlin.de</a>	yes	no	yes	Use and user statistics may be downloaded in CSV format.
Copenhagen <a href="https://www.opendata.dk/city-of-copenhagen">https://www.opendata.dk/city-of-copenhagen</a>	Not applicable	Not applicable	Not applicable	The former open data portal of the city of Copenhagen, <a href="https://data.kk.dk">https://data.kk.dk</a> , has been integrated into the national open data portal of Denmark.
Dublin <a href="https://data.smartdublin.ie">https://data.smartdublin.ie</a>	no	yes	no	
Florence <a href="https://opendata.comune.fi.it">https://opendata.comune.fi.it</a>	no	yes	no	Also proposed through the call to the open data community.
Gdansk <a href="https://www.gdansk.pl/en">https://www.gdansk.pl/en</a>	no	no	no	
Ghent <a href="https://data.stad.gent">https://data.stad.gent</a>	no	no	no	
Helsinki Region <a href="https://hri.fi/en_gb">https://hri.fi/en_gb</a>	yes	yes	no	Also proposed through the call to the open data community.
Lisbon <a href="https://lisboaaberta.cm-lisboa.pt/index.php/en">https://lisboaaberta.cm-lisboa.pt/index.php/en</a>	yes	yes	no	The range of available user statistics is limited.

City/Region	Reuse case section	Use(r) statistics	Report(s) on impact	Remarks
London <a href="https://opendata.london.ca">https://opendata.london.ca</a>	no	no	no	
Paris <a href="https://opendata.paris.fr">https://opendata.paris.fr</a>	yes	no	no	One statistic regarding dataset popularity by downloads is available on the homepage.
Stockholm <a href="https://dataportalen.stockholm.se/dataportalen">https://dataportalen.stockholm.se/dataportalen</a>	no	no	no	
Thessaloniki <a href="https://opendata.thessaloniki.gr/en">https://opendata.thessaloniki.gr/en</a>	no	no	no	
Vienna <a href="https://digitales.wien.gv.at/en">https://digitales.wien.gv.at/en</a>	yes	no	no	
Vilnius <a href="https://api.vilnius.lt">https://api.vilnius.lt</a>	no	no	no	
Madrid <a href="https://datos.madrid.es/portal/site/egob">https://datos.madrid.es/portal/site/egob</a>	yes	no	no	Proposed through the call to the open data community.
Bordeaux <a href="https://opendata.bordeaux-metropole.fr/pages/accueil">https://opendata.bordeaux-metropole.fr/pages/accueil</a>	yes	no	no	Proposed through the call to the open data community.
Eindhoven <a href="https://data.eindhoven.nl/pages/home">https://data.eindhoven.nl/pages/home</a>	yes	no	no	Proposed through the call to the open data community.
Zaragoza <a href="https://www.zaragoza.es/sede/portal/datos-abiertos/">https://www.zaragoza.es/sede/portal/datos-abiertos/</a>	yes	yes	no	Proposed through the call to the open data community.

## Annex IV– Impact measurement on national and local open data portals

Impact type/indicator name	Metric/question	Sources	Remarks
<b>CYPRUS</b>			
General/reuse description	All reuse cases include a short general description	Data reusers	The portal allows new reuse cases to be submitted for inclusion via a specific form.
General/reuse type	All reuse cases are tagged according to one of the following reuse types: mobile app, web app, desktop app, internet service, G2G app, iOS, android, web platform, dashboard, internal control system, website, Windows.	Data reusers	Fixed typology by portal provider
General/reuse theme	All reuse cases are tagged according to one of the following themes: government and public sector; transportation; provinces, municipalities and zones; energy; justice, legal system and public safety; health; environment; education, culture and sports; population, society and employment; economy and finance; agriculture, fisheries, forestry and food; science and technology.	Portal team	Fixed typology by portal provider
General/detailed cases	Selected reuse cases are described in more detail.	Additional research by portal team	Presented in a dedicated articles section
General/links to datasets	Links to source datasets are provided for some highlighted reuse cases.	Data reusers	Presented in a dedicated articles section
General/number of reuse cases	Cumulative number of featured reuse cases since mid-2016.	Portal statistics	It is unclear whether this tracking is done automatically.
General/requests for data	The portal tracks number of requests submitted for data since mid-2016.	Portal statistics	Relevant for estimated, not realised impact
General/dataset ranking	The portal provides a ranking of datasets based on the number of user visits.	Portal statistics	Somewhat relevant for estimated, not realised impact

Impact type/indicator name	Metric/question	Sources	Remarks
Economic/use of open data	<ul style="list-style-type: none"> <li>- State whether your organisation/business owns/publishes/updates/uses open data and/or further open data (i.e. is a user of public information or has information in the market).</li> <li>- To what extent would you say that you use open public sector data for the operation of your business/organisation?</li> <li>- Please indicate the categories or domains of open data that you or your organisation (organisation, business, link, etc.) have/have used or are currently using.</li> <li>- What percentage of the data used by your organisation/business is classified in the open data category?</li> <li>- For which of the following activities/actions do you use open data: research and development / creation of new products/services, development/creation of new applications, to improve existing products/services, optimisation of internal processes, to inform the public/users of services etc., data as a service, other, I do not know.</li> <li>- State whether your organisation/business produces data itself.</li> </ul>	Survey of data reusers	12 categories of open data were provided as choices for the third indicator; these are the same categorisations applied to reuse cases.
Economic/created jobs	<ul style="list-style-type: none"> <li>- Would you say that new jobs have been created (directly or indirectly) in your company/organisation due to the possibility of exploitation/reuse of open data?</li> <li>- If so, how many new posts have been created in recent years?</li> </ul>	Survey of data reusers	
Economic/future jobs	<ul style="list-style-type: none"> <li>- Does your organisation/company plan to hire additional people who will deal either directly or indirectly with the use/reuse of open data in the next 2 years?</li> <li>- If so, how many new jobs are expected to be created?</li> </ul>	Survey of data reusers	This indicator refers to estimated impact.
Economic/general turnover	<ul style="list-style-type: none"> <li>- What is the range of the average annual capital turnover of your organisation/business for the last 5 years (or if the company is less than 5 years old, since its inception)?</li> <li>- Do you believe that the availability or lack of availability of open public data affects the turnover of your organisation/business?</li> </ul>	Survey of data reusers	



Impact type/indicator name	Metric/question	Sources	Remarks
Economic/direct turnover	- What percentage of your turnover for 2020 (or 2019 if the 2020 data is not yet available) do you estimate is directly related to the reuse of open data resources?	Survey of data reusers	
Economic/indirect turnover	- What percentage of your turnover for 2020 (or 2019 if the 2020 data is not yet available) you estimate is associated with indirect benefits from the reuse of open data resources: performance benefits, production benefits, time benefits?	Survey of data reusers	
Economic/estimated turnover	- Do you think that having more data would positively affect the turnover of your organisation/business? - If 'Yes', what is the projected annual growth (%) of turnover associated with open data for the next 2 years?	Survey of data reusers	This indicator refers to estimated impact.
General/benefits of using open data	- What, in your opinion, are the main advantages/benefits arising from making open data available for further use? Enhancing transparency and accountability of the public sector; reducing operating costs of the public sector; improving the quality of state services; facilitating citizens' interaction with government; encouraging entrepreneurship; creating new businesses (including start-ups); encouraging research, development, and innovation; improving the decision-making processes of public bodies; other; I do not know. - Indicate how important each advantage you have chosen in the above question is.	Survey of data reusers	Survey participants were asked to assign each advantage a number between one and five, where one was 'not at all', four was 'very important' and five was 'I do not know.' This indicator refers to estimated impact.
Economic/business benefits	- What, in your opinion, are the general benefits of exploiting/reusing open data for an enterprise/organisation? Production/creation of new products, applications, or services; enhancing knowledge for specific sectors/areas (e.g. tourism); improved business activities; improving competitiveness; reduction of resources needed to conduct various studies (e.g. market analysis); increased revenue/sales; possibility of new studies; other; I do not know.	Survey of data reusers	This indicator refers to estimated impact.

Impact type/indicator name	Metric/question	Sources	Remarks
Social/social groups	<ul style="list-style-type: none"> <li>- Based on your experience, do you think there are any specific social groups that benefit from the use of open data and its availability for exploitation, i.e. reuse?</li> <li>- If so, which social groups do you think benefit from it? Elderly people; people with disabilities or other special needs; refugees and/or asylum seekers; immigrants (third-country nationals); children and adolescents; victims of violence or exploitation; other.</li> </ul>	Survey of data reusers	<p>For the second indicator, participants were asked to assign each social group a number between one and five, where one was 'do not benefit at all', four was 'benefit very much' and five was 'I do not know.'</p> <p>This indicator refers to estimated impact.</p>
Social and environmental/groups	<ul style="list-style-type: none"> <li>- In general, to what extent do you consider open data to have helped the following: inclusion of marginalised/vulnerable groups in society; raising public awareness about housing opportunities and the real estate market; meeting social challenges; citizens' perception of water and air quality; citizens' perception of the various forms of pollution; promoting and supporting the use of environmentally friendly means of transport; better waste management; use of alternative energy sources.</li> </ul>	Survey of data reusers	
General/impact evaluation	<ul style="list-style-type: none"> <li>- Finally, how do you assess the impact of open data on the Cypriot economy and society? Very negative; negative; neither negative nor positive; positive; very positive; I do not know.</li> </ul>	Survey of data reusers	
<b>GERMANY</b>			
General/number of reuse cases	Cumulative number of featured reuse cases	Portal statistics	
General/reuse description	All reuse cases include a short general description	Data reusers or portal team	The portal allows new reuse cases to be submitted for inclusion via a general contact form.

Impact type/indicator name	Metric/question	Sources	Remarks
General/reuse type	All reuse cases are tagged according to one of the following reuse types: website, visualisation, tool, concept, mobile app, scientific publication, miscellaneous.	Data reusers or portal team	Fixed typology by portal provider
General/reuse theme	All reuse cases are tagged according to one of the following themes: population and society; education, culture and sport; health; agriculture, fisheries, forestry and food; government and public sector; regions and cities; environment; traffic; economy and finance; science and technology.	Data reusers or portal team	Fixed typology by portal provider
General/reuse platform	All reuse cases are tagged according to one of the following systems: web, Android, iOS, other.	Data reusers or portal team	Fixed typology by portal provider
General/reuse keywords	Most reuse cases are tagged with keywords such as 'bicycle', 'climate change' or 'water'.	Data reusers or portal team	
General/links to datasets	Links to source datasets are provided for some reuse cases.	Data reusers or portal team	
<b>ESTONIA</b>			
General/reuse description	All reuse cases include a short general description	Portal team or data reusers	
General/dataset ranking	The portal provides a ranking of datasets based on the number of user downloads in the past 30 days.	Portal statistics	Relevant for estimated, not realised impact
General/keyword ranking	The portal provides a ranking of the 100 keywords most frequently searched for by portal visitors over the last 28 days.	Portal statistics	Somewhat relevant for estimated, not realised impact

Impact type/indicator name	Metric/question	Sources	Remarks
General/reuser groups	Who are the main reusers of open data (e.g. universities, students, start-ups) and who could be reusers?	Surveys of public sector agencies	Surveys were conducted in 2020 and 2021. Both consisted of seven open-ended questions, but the exact wording of the 2021 questionnaire is not included in the published report. The report's conclusions imply that the questions asked were similar or identical. The first part of the question is relevant for realised impact, whereas the second part refers to estimated impact.
General/support to reusers	<ul style="list-style-type: none"> <li>- Why have you made the data available as open data? Is it possible for those interested in reusing data to request access to data that is not available today?</li> <li>- How do you present existing data to potential stakeholders (e.g. information day, press releases, publication on a website or information portal)?</li> <li>- How satisfied are users with the open data published by your organisation?</li> <li>- Does your organisation have the ability to provide support to consumers of open data (consulting, user manuals)? If not, why would you need support/assistance to achieve this capability?</li> </ul>	Survey of public sector agencies	These indicators can be considered as moderating variables affecting the impact of data publication.
Public sector impact/government benefits	What benefits have you received from the disclosure of open data (e.g. reduced number of requests for information and response times, reuse of data, positive image)?	Survey of public sector agencies	
General/portal quality	Visitors are asked to rank the portal's quality on a ten-point scale and given the option to provide additional written feedback.	Portal feedback button/pop-up	This indicators can be considered as a moderating variable affecting the impact of data publication.

Impact type/indicator name	Metric/question	Sources	Remarks
<b>SPAIN</b>			
General/reusers	The total number of reuse companies .	Portal statistics	The portal tracks the number of known reusing companies since January 2017.
General/reuser description	All reuse companies are briefly profiled.	Data reusers	New companies may submit themselves for inclusion via a specific form.
General/reuser theme	All reuse companies are tagged according to one or more of the following themes: science and technology, economy, environment, public sector, rural environment, society and well-being, urban planning and infrastructure, housing, culture and leisure, transportation, industry; tourism, demographics, trade, education, employment, health, sport, energy, legislation and justice, security.	Data reusers	Fixed typology by portal provider
General/reuser age	All reuse companies are tagged according to the amount of time they have been established: less than 5 years, from 5 to 10 years, from 10 to 20 years, more than 20 years.	Data reusers	Fixed typology by portal provider
General/reuser keywords	Most reusing companies are tagged with keywords such as 'cartography', 'smart farming' or 'winemaking'.	Portal team	
General/number of reuse cases	Cumulative number of featured reuse cases	Portal statistics	The portal tracks the number of known reuse cases since March 2013
General/reuse description	All reuse cases include a short general description.	Data reusers	The portal allows new reuse cases to be submitted for inclusion via a specific form
General/reuse theme	All reuse cases are tagged according to the same themes used for reusing companies.	Data reusers	Fixed typology by portal provider
General/reuse platform	All reuse cases are tagged according to one or more of the following platforms: website, Android, iOS, desk, mobile web, Windows, MacOS, mobile (native), Linux, Symbian, other.	Data reusers	Fixed typology by portal provider

Impact type/indicator name	Metric/question	Sources	Remarks
General/reuse keywords	Most reuse cases are tagged with keywords such as 'Madrid', 'geospatial data' or 'beaches'.	Portal team	
General/links to datasets	Case subpages consistently contain links to source datasets.	Initiative leaders	Fixed typology by portal provider
General/highlighted cases	Certain reuse cases are highlighted separately and described in greater detail.	Additional research by portal team	Presented in separate news, newsletters and blog sections
General/dataset ranking	The portal provides a ranking of the 10 most-visited datasets since 1 January 2016 with the option to see the most-visited datasets in any month within that date range.	Portal statistics	Somewhat relevant for estimated, not realised impact
<b>FINLAND</b>			
General/number of reuse cases	Cumulative number of featured reuse cases	Portal statistics	
General/reuse description	All reuse cases include a short general description.	Data reusers	The portal allows new reuse cases to be submitted for inclusion via a specific form
General/reuse theme	All reuse cases are tagged according to one or more of the following themes: regions and cities; transportation; environment and nature; population and society; economy and finance; government and public sector; built environment and infrastructure; culture, art and leisure; health; travel and tourism; education and sports; science and technology.	Portal team	Fixed typology by portal provider
General/reuse platform	All reuse cases are tagged according to one of the following platforms: Android, iOS Apple, Windows, Mac OS X, other.	Portal team	Fixed typology by portal provider
General/reuse keywords	Most reuse cases are tagged with keywords such as 'visualisation', 'Helsinki' or 'linked data'.	Portal team	
General/links to datasets	Case subpages consistently contain links to source datasets.	Data reusers	
General/data by administrative sector	The portal tracks the distribution of data uploads across 11 administrative sectors and 31 subsectors.	Portal statistics	Measures data publication, not impact, but could prove useful to track reuse across sectors.
General/downloads	The portal provides a ranking of the nine datasets with the most cumulative downloads.	Portal statistics	

Impact type/indicator name	Metric/question	Sources	Remarks
General/page visits	The portal provides rankings of the 20 datasets with the most page visits in the past week, month or year.	Portal statistics	
General/keyword ranking	The portal provides a ranking of the 20 keywords most frequently searched for by portal visitors over the last week, month or year.	Portal statistics	Somewhat relevant for estimated, not realised impact
<b>FRANCE</b>			
General/number of reuse cases	Cumulative number of featured reuse cases	Portal statistics	
General/reuse description	All reuse cases include a short general description	Data reusers	The portal allows new reuse cases to be submitted for inclusion via a specific form
General/reuse theme	All reuse cases are tagged according to one or more of the following themes: health, transport and mobility, land use planning and housing, food and agriculture, culture and leisure, economy and business, environment and energy, employment and training, politics and public life, security, education and research, society and demography, law and justice, open data tools, other.	Data reusers	Fixed typology by portal provider
General/reuse type	All reuse cases are tagged according to one of the following types: application, visualisation, API, paper, blog article, press article, idea.	Data reusers	Fixed typology by portal provider
General/links to datasets	Case subpages consistently contain links to source datasets.	Data reusers	
General/dataset ranking	The portal provides rankings of datasets based on the number of downloads and unique downloads.	Portal statistics	
<b>CROATIA</b>			
General/reuse description	All reuse cases include a short general description.	Data reusers or portal team	The portal allows new reuse cases to be submitted by logged-in users
General/reuse fee	All reuse cases are tagged as being either free or paid.	Data reusers or portal team	Fixed typology by portal provider

Impact type/indicator name	Metric/question	Sources	Remarks
General/highlighted cases	Certain reuse cases are highlighted separately.	Portal team	Presented in a separate section labelled 'Examples of good practice'
General/reuse additions by type	The portal tracks added reuse cases by type since December 2021.	Portal statistics	No new cases have been added; all are applications
General/dataset suggestions	The portal tracks new and resolved suggestions for datasets since January 2022.	Portal statistics	
<b>IRELAND</b>			
General/number of reuse cases	Cumulative number of featured reuse cases	Portal statistics	
General/reuse description	All reuse cases include a short general description.	Data reusers	The portal allows new reuse cases to be submitted for inclusion via a specific form
General/reuse keywords	Most reuse cases are tagged with keywords such as 'archaeology', 'transport' or 'marine'.	Data reusers	Keywords are suggested by submitters of reuse cases
General/highlighted cases	Certain reuse cases are highlighted separately.	Selection by portal team	Presented in a separate 'Impact stories' section
General/requests for data	The portal tracks number of requests submitted for data and status of requests between August 2016 and September 2021.	Portal statistics	Relevant for estimated, not realised impact
General/dataset ranking	The portal provides a ranking of the top 10 most-viewed datasets.	Portal statistics	Somewhat relevant for estimated, not realised impact
General/dataset ranking	The portal provides a ranking of the top 10 most-downloaded datasets.	Portal statistics	
General/keyword ranking	The portal provides a ranking of the 10 keywords most frequently searched for by portal visitors.	Portal statistics	Somewhat relevant for estimated, not realised impact
<b>LITHUANIA</b>			
General/number of reuse cases	Cumulative number of featured reuse cases	Portal statistics	



Impact type/indicator name	Metric/question	Sources	Remarks
General/reuse type	Cases are subdivided into 'examples of data use' and 'developed applications'.	Data reusers or portal team	
General/reuse description	All reuse cases include a short general description.	Data reusers or portal team	The portal allows new reuse cases to be submitted by logged-in users
General/reuse benefits	The benefits of some reuse cases are specified without a predefined categorisation.	Data reusers or portal team	
General/reuse beneficiaries	The beneficiaries of some reuse cases are specified without a predefined categorisation.	Data reusers or portal team	
General/links to datasets	Some case subpages contain links to source datasets.	Data reusers or portal team	
General/additions to data by theme	New dataset additions in 2021 were tracked in each of the following themes: economic activity and business, economics and finance, environment, state and public administration, transport and communication, healthcare, social security, education.	2021 Review of Public Sector Data Opening (Belickas, 2022)	Measures data publication, not impact, but could prove useful to track reuse across categories.
<b>NETHERLANDS</b>			
General/number of reuse cases	Cumulative number of featured reuse cases	Portal statistics	
General/reuse description	All reuse cases include a short general description.	Data reusers	The portal allows new reuse cases to be submitted for inclusion via a specific form
General/reuse theme	All reuse cases are tagged according to one of the following themes: nature and environment, traffic, care and health, management, education and science, economy, space and infrastructure, housing, culture and recreation, finance, public order and safety, agriculture, migration and integration, social security. 10 of these themes have between one and four subthemes for further filtering.	Data reusers	Fixed typology by portal provider
General/dataset ranking	The portal provides rankings of the top 10 most-viewed datasets each year from 2016 to 2021.	Portal statistics	Somewhat relevant for estimated, not realised impact

Impact type/indicator name	Metric/question	Sources	Remarks
General/keyword ranking	The portal provides rankings of the 10 keywords most frequently searched for by portal visitors in 2019, 2020 and 2021.	Portal statistics	Somewhat relevant for estimated, not realised impact
General/dataset prioritisation	'High-value' datasets are identified by their contributions to transparency, legal obligation and cost reduction, as well as its target audience and reuse potential.	Municipal and provincial high-value data lists Additional research by portal team	Relevant for the measurement of estimated impact.
General/impact stories	Cases of political, social, environmental and economic impact are highlighted through impact stories	Interviews conducted by portal team; Data reusers	Presented in a separate 'Impact stories' section
General/data communities	Data communities consolidate open data, data quality assessments, data publishers and reusers and discussion forums for the following domains: energy, mobility, social security, education, migration.	Portal team, portal users, data publishers and data reusers	
<b>POLAND</b>			
General/reuse cases	The total number of featured reuse cases	Portal statistics	
General/reuse description	All reuse cases include a short general description.	Data reusers	The portal allows new reuse cases to be submitted for inclusion via a specific form
General/reuse platform	All reuse cases are tagged according to one of the following platforms: website, application, other.	Data reusers	
General/reuse keywords	Most reuse cases are tagged with keywords such as 'kindergartens', 'search engine' or 'credibility of companies'.	Data reusers	
General/highlighted cases	Certain reuse cases are highlighted separately and described.	Additional research by portal team	Presented in a separate 'News' section
<b>UKRAINE</b>			
General/highlighted cases	Certain reuse cases are highlighted separately in a dedicated news section. Their descriptions may include the nature of the service, potential benefits and intended beneficiaries.	Interviews with reusers Additional research by portal team	Presented under the 'News' section

Impact type/indicator name	Metric/question	Sources	Remarks
General/popular datasets	Unknown	Portal statistics	On the archived version of the portal it was not possible to assess whether this indicator is measured based on visits or downloads.
General/Impact categories	The portal compiles relevant datasets and examples of data reuse on subpages devoted to the impacts of open data related to construction, infrastructure, health, ecology, business, the legal branch, local government, state supervision, financial transparency, forestry and the property and income declarations of public officials.	Portal team	
General/sectoral benefits	What are the potential benefits of open data in the areas of state supervision, financial transparency, forestry, construction, roads, healthcare, ecology, business and local government?	Written and oral interviews with NGO representatives, public and private sector data reusers, state officials and other relevant personnel	The Ministry of Digital Transformation has conducted nine impact studies across different sectoral areas.  Concerns estimated impact
General/paths to sectoral impacts	How do reusers in the public and private sectors currently apply data towards specified sectoral goals in the areas of state supervision, financial transparency, forestry, construction, roads, healthcare, ecology, business and local government?	Written and oral interviews with NGO representatives, public and private sector data reusers, state officials and other relevant personnel	The Ministry of Digital Transformation has conducted nine impact studies across different sectoral areas.
General/portal accessibility	Technical aspects, e.g. API query limitations	Internal evaluation	
<b>BARCELONA</b>			
General/reuse description	All reuse cases include a short general description.	Data reusers	The portal allows new reuse cases to be submitted for inclusion via a specific form
General/links to datasets	Links to source datasets are provided for most reuse cases	Data reusers	

Impact type/indicator name	Metric/question	Sources	Remarks
General/reuse theme	Visualisations under the following themes may be consulted: air quality, trees, biking stations and traffic accidents managed by the Guàrdia Urbana de Barcelona.	Portal team	Fixed typology by portal provider
General/dataset ranking	The portal provides rankings of the top five most-visited datasets both over the last 15 days and cumulatively since February 2017.	Portal statistics	Somewhat relevant for estimated, not realised impact
General/dataset ranking	The portal provides rankings of the top five most-downloaded datasets by IP address both over the last 15 days and cumulatively since February 2017	Portal statistics	Relevant for estimated, not realised impact
General/dataset ranking	The portal provides rankings of the top five most-downloaded datasets by absolute value both over the last 15 days and cumulatively since February 2017.	Portal statistics	Relevant for estimated, not realised impact
Impact visibility	Cases of academic, professional and administrative open data use are briefly profiled.	Additional research/compiling by portal team	Presented under separate 'Usage stories' section
<b>BERLIN</b>			
General/reuse cases	The total number of featured reuse cases	Portal statistics	
General/reuse description	All reuse cases include a short general description	Data reusers or portal team	The portal allows new reuse cases to be submitted for inclusion via a general contact form
General/reuse keywords	All reuse cases are tagged with keywords such as 'Christmas', 'bicycles' or 'app'.	Data reusers or portal team	
General/links to datasets	Case subpages contain links to source datasets.	Data reusers or portal team	
General/highlighted datasets	Frequently reused datasets are highlighted separately and described.	Additional research by portal team	Presented under separate 'Interaction' section

<p>Open data survey</p>	<ul style="list-style-type: none"> <li>- I count myself as a stakeholder in this area: administration; digital civil society/digital volunteering; economy/business; science; other.</li> <li>- I am active in: Berlin; Berlin area; outside of Berlin/Brandenburg.</li> <li>- I am familiar with open data: not at all; a little bit; very familiar.</li> <li>- The importance of open data in Berlin will increase now and in the future.</li> <li>- These goals of using open data are most important to me: transparency of administrative actions; basis for possible new business models; increased efficiency of administrative work; automation of administrative activities; other</li> <li>- Considering the relevance of open data, the topic enjoys an appropriate level of importance in the state of Berlin.</li> <li>- The topic of open data should be prioritised differently: higher; lower; the prioritisation is good; I do not know.</li> <li>- I have a good understanding of the data that the state of Berlin already provides.</li> <li>- I have already used data from the state of Berlin or the districts (e.g. for my own projects): yes; no; no answer.</li> <li>- The data I would like to use is published on the data portal: yes; no; partially; other.</li> <li>- Are there certain subject areas from which you think significantly more open data should be published? (open-ended)</li> <li>- Open data is currently primarily used by this group(s): administration; digital civil society/digital volunteering; economy/business; science; none; I do not know; other.</li> <li>- In the future, I see great potential for using open data in this group(s): administration; digital civil society/digital volunteering; economy/business; science; none; I do not know; other.</li> <li>- The data is usually published in such a way that it is easily accessible to me.</li> <li>- The data is usually complete and of high quality.</li> <li>- The data is usually published in a format that is easy to use (e.g. CSV, Excel).</li> <li>- The conditions for reuse of data (e.g. licences) are usually clearly and</li> </ul>	<p>Open Data Strategy Berlin survey</p>	<p>For questions with ‘other’ options, visitors were asked to specify if they selected ‘other.’</p> <p>For non-open-ended questions, if different response options are not specified, respondents were asked to select a response from the following scale:</p> <ol style="list-style-type: none"> <li>1. I do not agree;</li> <li>2. I somewhat disagree;</li> <li>3. I do not know;</li> <li>4. I somewhat agree;</li> <li>5. I totally agree;</li> <li>6. No answer.</li> </ol>
-------------------------	--	---	--

Impact type/indicator name	Metric/question	Sources	Remarks
	<p>adequately marked.</p> <ul style="list-style-type: none"> <li>- The data can usually be called up automatically (e.g. via the interface) in a reliable manner.</li> <li>- The preparation of open data for subsequent use is very time-consuming.</li> <li>- There is sufficient expertise in my area for the processing of open data.</li> <li>- From my point of view, the biggest obstacle to the provision of open data in Berlin is: (open-ended).</li> <li>- I can imagine providing data myself: yes; no; I already do this.</li> <li>- I would like to learn more about open data and I think cross-departmental exchange is good: yes; no; I do not know.</li> <li>- The following is important to me for the current open data strategy process: (open-ended).</li> </ul>		
<b>DUBLIN</b>			
General/dataset ranking	The portal provides a ranking of the top 10 datasets based on the number of user visits.	Portal statistics	Relevant for estimated, not realised impact
General/keyword ranking	The portal provides a ranking of the 10 keywords most frequently searched for by portal visitors.	Portal statistics	Somewhat relevant for estimated, not realised impact
<b>FLORENCE</b>			
General/dataset ranking	The portal provides rankings of the top 10 most-downloaded datasets, both cumulatively and for every month since December 2018.	Portal statistics	Relevant for estimated, not realised impact
<b>HELSINKI REGION</b>			
General/reuse cases	The total number of featured reuse cases	Portal statistics	
General/reuse description	All reuse cases include a short general description.	Data reusers	The portal allows new reuse cases to be submitted for inclusion via a specific form
General/reuse type	All reuse cases are tagged according to one of the following reuse types: visualisations, applications, data journalism, examples from elsewhere, tools, objects, city models, data, instructions.	Portal team	Fixed typology by portal provider

Impact type/indicator name	Metric/question	Sources	Remarks
General/reuse platform	All reuse cases are tagged according to one of the following platforms: Android, iOS, iOS, Windows phone, web, iPhone, BlackBerry, Pebble OS, Symbian, Minecraft, J2ME, Windows, Linux.	Data reuser	Fixed typology by portal provider  The second iOS tag appears to be redundant.
General/reuse keywords	All reuse cases are tagged with keywords such as 'trains', 'traffic' or 'weather'.	Portal team	
General/links to datasets	Links to source datasets are provided for all reuse cases.	Data reuser	
General/reuse suggestions	The portal allows suggestions for reuse applications to be submitted through a separate form and lists these on a subpage.	Portal visitors	
General/dataset ranking	The portal provides rankings of the top 15 most-visited datasets cumulatively or in any period between 2020 and the present.	Portal statistics	Somewhat relevant for estimated, not realised impact
General/category distribution of reuse	The portal tracks the distribution of reuse cases by type.	Portal statistics	
<b>LISBON</b>			
General/reuse description	All reuse cases include a short general description.	Portal team or data reusers  Interviews conducted at Smart Open Lisboa 2019	
General/dataset ranking	The portal provides a ranking of the top five most-visited datasets.	Portal statistics	Somewhat relevant for estimated, not realised impact
General/dataset ranking	The portal provides a ranking of the top five most-downloaded datasets.	Portal statistics	Relevant for estimated, not realised impact
<b>PARIS</b>			
General/reuse description	All reuse cases include a short general description.	Data reuser	

Impact type/indicator name	Metric/question	Sources	Remarks
General/links to datasets	Links to source datasets are provided for all reuse cases.	Portal design	The portal allows new reuse cases to be submitted through the subpages for individual datasets
General/links to cases	Dataset subpages link to reuse cases created with that data	Portal design	
General/dataset ranking	The portal provides a ranking of the top five most-downloaded datasets.	Portal statistics	Relevant for estimated, not realised impact
<b>VIENNA</b>			
General/reuses	The total number of reuse cases	Portal statistics	
General/reuse description	All reuse cases include a short general description.	Data reuser	The portal allows new reuse cases to be submitted for inclusion via a specific form
General/reuse type	All reuse cases are tagged according to one of the following reuse types: data preparation tool, concept, mobile app, tool, tutorial, visualisation, web app, scientific publication, participation.	Data reuser	Fixed typology by portal provider
General/reuse platform	All reuse cases are tagged according to one of the following platforms: Android, BlackBerry, browser www, Firefox OS, iOS, Linux, MacOSX, OSX, Sonstiges, Symbian, Watch OS, WebOS, Windows, Windows 8, Windows Phone.	Data reuser	Fixed typology by portal provider
General/reuse theme	All reuse cases are tagged according to one of the following themes: work, population, education and research, finance and accounting, geography and planning, society and social affairs, health, art and culture, agriculture and forestry, sport and leisure, environment, traffic and technology, administration and politics, economy and tourism.	Portal team	Fixed typology by portal provider
General/reuse data source	All reuse cases are tagged with the producer of their source data. Example producers include 'City of Innsbruck' and 'Federal Geological Survey'.	Portal team	
General/links to datasets	Links to source datasets are provided for all reuse cases.	Data reuser	
General/category distribution of reuse	The portal tracks the distribution of reuse cases by type.	Portal statistics	



Impact type/indicator name	Metric/question	Sources	Remarks
General/category distribution of reuse	The portal tracks the distribution of reuse cases by source of data.	Portal statistics	
<b>MADRID</b>			
General/reuses	The total number of reuse cases	Portal statistics	
General/reuse description	All reuse cases include a short general description.	Data reusers	The portal allows new reuse cases to be submitted for inclusion via a specific form
<b>BORDEAUX</b>			
General/reuse description	All reuse cases include a short general description.	Data reusers	The portal allows new reuse cases to be submitted for inclusion via a specific form
General/reuse theme	All reuse cases are tagged according to one of the following themes: mobility, society, citizenship, territory.	Portal team	Fixed typology by portal provider
General/reuse theme	All reuse cases are tagged with the theme of their source dataset: citizenship and administration; displacement; reference mapping; roads/public spaces; sustainable development; urban planning-layout; culture, sports and leisure; living environment; economic development; sustainable development (old); picto test – citizenship and administration.	Portal team	Fixed typology by portal provider The final two themes appear to be redundant.
General/reuse type	Visitors may choose to see only applications or visualisations.	Portal team	Fixed typology by portal provider
Data reuse discussion	The portal's forum contains a board for discussion and promotion of reuse cases.	Portal visitors Data reusers	All posts thus far were made by reusers.
<b>EINDHOVEN</b>			
General/reuse description	All reuse cases include a short general description.	Portal team	All reuse cases are graphics created by the portal team

Impact type/indicator name	Metric/question	Sources	Remarks
General/reuse theme	All reuse cases are tagged according to one of the following themes: population & governance (population, public service); economy, work & income (work); physical living environment (environment & sustainability, public space & greenery, traffic & transport, living); social living environment (education; safety; leisure, culture & sport; well-being & health).	Portal team	Fixed typology by portal provider Themes in parentheses are subthemes under a broader heading
General/dataset ranking	The portal provides a ranking of the top five most-downloaded datasets.	Portal statistics	Relevant for estimated, not realised impact
<b>ZARAGOZA</b>			
General/reuse description	All reuse cases include a short general description.	Data reusers or portal team	The portal allows new reuse cases to be submitted by logged-in users
General/reuse type	All reuse cases are tagged according to one of the following platforms: Android, iOS, web, Windows, Linux.	Data reusers or portal team	Fixed typology by portal provider
General/reuse device	All reuse cases are tagged according to one of the following device types: mobile, tablet, desktop, tablet.	Data reusers or portal team	Fixed typology by portal provider 'Tablet' appears twice; this seems to be a redundancy.
General/reuse keywords	Users may filter reuse cases by keywords, including 'mobility', 'music' and 'kids'.	Data reusers or portal team	
General/data format	The portal tracks the distribution of downloads across the following formats: JSON, GeoJSON, CSV, solr.xml, SolJSon, XML, PFD, GeoRSS, N3, RSS, JSON-LD, RDF, TTL, XLS, DOC, KML, ICS. This is expressed in both actual value and percentage form.	Portal statistics	
General/retrieval method	The portal tracks the distribution of downloads across the following HTML retrieval methods: GET, POST, PUT, DELETE, HEAD. This is expressed in both actual value and percentage form.	Portal statistics	

## Annex V – open data intermediary case study overview

Name <i>Source article link</i>	Type of organisation	Reusers involved	Service provided	Relation with other intermediaries	Trust mechanism	Impact domains
Centre for Higher Education Trust (CHET) <a href="https://doi.org/10.6084/m9.figshare.1449222.v1">https://doi.org/10.6084/m9.figshare.1449222.v1</a>	Non-governmental organisation	<ul style="list-style-type: none"> <li>- University planners and councils at South African public universities</li> <li>- Researchers studying university performance</li> </ul>	<ul style="list-style-type: none"> <li>- Republishes data in tables corresponding to specific performance indicators</li> <li>- Maintains interface for generating custom graphs and data tables and comparing university performance</li> </ul>	Reformats and republishes data collected by the Department of Higher Education and Training (DHET)	<ul style="list-style-type: none"> <li>- History of interactions with DHET</li> <li>- Social capital accrued among universities – performance indicators had been established for 16 years at time of study</li> <li>- Superior technical capital – researchers and universities report greater ease of use for CHET tables</li> </ul>	<p>Social (educational)</p> <p>Political (possible cost savings for DHET)</p>
ISDC Software <a href="https://doi.org/10.6084/m9.figshare.1449222.v1">https://doi.org/10.6084/m9.figshare.1449222.v1</a>	Private company	22 South African public universities	<ul style="list-style-type: none"> <li>- Offers a tool for accessing updated data on higher education</li> <li>- Processes, repackages and stores data</li> <li>- Maintains its own portal for data access</li> </ul>	Reformats and republishes data collected by DHET	<ul style="list-style-type: none"> <li>- Economic capital from licensing modules/institutional partnerships</li> <li>- Some cultural/social capital in academic sector (established partnership with University of Johannesburg)</li> <li>- Technical capital</li> </ul>	<p>Social (educational)</p> <p>Political (possible cost savings for DHET)</p>
Budgit <a href="https://figshare.com/articles/journal_contribution/Open_Data_In_intermediaries_in_Developing_Countries/1449222/1">https://figshare.com/articles/journal_contribution/Open_Data_In_intermediaries_in_Developing_Countries/1449222/1</a>	Non-governmental organisation	<ul style="list-style-type: none"> <li>- Citizens</li> <li>- Private enterprises</li> </ul>	<ul style="list-style-type: none"> <li>- Aggregates and republishes open data related to the public budget of Nigeria</li> <li>- Sells infographics to private-sector companies</li> </ul>	None recorded	<ul style="list-style-type: none"> <li>- Technical capital</li> <li>- Economic capital from donor funding and sale of infographics</li> </ul>	Political (encourages government transparency)

Name <i>Source article link</i>	Type of organisation	Reusers involved	Service provided	Relation with other intermediaries	Trust mechanism	Impact domains
Indianpetro <a href="https://figshare.com/articles/journal_contribution/Open_Data_Intermediaries_in_Developing_Countries/1449222/1">https://figshare.com/articles/journal_contribution/Open_Data_Intermediaries_in_Developing_Countries/1449222/1</a>	Private company	<ul style="list-style-type: none"> <li>- Citizens</li> <li>- Private enterprises</li> <li>- Researchers studying energy and infrastructure</li> </ul>	<ul style="list-style-type: none"> <li>- Aggregates and reports data from the energy and infrastructure sectors in India</li> <li>- Combines open and private data</li> <li>- Incorporates data from other national administrations</li> <li>- Organises conferences and forums on energy issues</li> </ul>	None recorded	Deploys privately raised economic capital to access data behind paywalls	<p>Political (encourages government transparency)</p> <p>Environmental (if data is applied in service of research or pro-environment lobbying)</p>
Open Knowledge Foundation Nepal <a href="https://figshare.com/articles/journal_contribution/Open_Data_Intermediaries_in_Developing_Countries/1449222/1">https://figshare.com/articles/journal_contribution/Open_Data_Intermediaries_in_Developing_Countries/1449222/1</a>	Non-profit network	Citizens	<ul style="list-style-type: none"> <li>- Republishes Nepalese open budget data on own website</li> <li>- Provides training on accessing open data</li> <li>- Hosts open data hacking meetups</li> </ul>	None recorded	Technical capital (internet access is a barrier to citizen data access in Nepal)	Political (encourages government transparency)

Name <i>Source article link</i>	Type of organisation	Reusers involved	Service provided	Relation with other intermediaries	Trust mechanism	Impact domains
World Bank Institute <a href="https://figshare.com/articles/journal_contribution/Open_Data_In_intermediaries_in_Developing_Countries/1449222/1">https://figshare.com/articles/journal_contribution/Open_Data_In_intermediaries_in_Developing_Countries/1449222/1</a>	International non-governmental organisation	<ul style="list-style-type: none"> <li>- Citizens</li> <li>- Journalists</li> <li>- Civil society actors</li> <li>- Technologists</li> </ul>	<ul style="list-style-type: none"> <li>- Aggregates and maps local development data from village development committees and national administration</li> <li>- Provides training on accessing open data</li> </ul>	None recorded	Technical capital (internet access is a barrier to citizen data access in Nepal)	<p>Political (encourages government transparency)</p> <p>Social (aims to reduce incidence of poverty in Nepal)</p>
DataMeet <a href="https://figshare.com/articles/journal_contribution/Open_Data_In_intermediaries_in_Developing_Countries/1449222/1">https://figshare.com/articles/journal_contribution/Open_Data_In_intermediaries_in_Developing_Countries/1449222/1</a>	Non-governmental organisation	<ul style="list-style-type: none"> <li>- Citizens</li> <li>- Researchers</li> </ul>	<ul style="list-style-type: none"> <li>- Volunteers extract and clean open data and republish sanitised, reorganised datasets via GitHub</li> <li>- Shares knowledge regarding data access</li> <li>- Combines government and non-commercial data</li> <li>- Liaises with Government of India Non-Commercial Sources to improve quality of data hosted on government portal</li> </ul>	None recorded	<ul style="list-style-type: none"> <li>- Technical capital</li> <li>- Social/cultural capital: relationship with India's governmental non-commercial sources</li> </ul>	<p>Political (encourages government transparency)</p> <p>Social (democratises data access)</p>

Name <i>Source article link</i>	Type of organisation	Reusers involved	Service provided	Relation with other intermediaries	Trust mechanism	Impact domains
Johns Hopkins University <a href="https://dl.acm.org/doi/pdf/10.1145/3511102">https://dl.acm.org/doi/pdf/10.1145/3511102</a>	Educational institution (university)	<ul style="list-style-type: none"> <li>- Citizens</li> <li>- Researchers studying the pandemic</li> <li>- Journalists</li> <li>- Regional and national administrations</li> </ul>	<ul style="list-style-type: none"> <li>- Aggregates and republishes COVID-19-related data from national and local health departments and governments on GitHub and ArcGIS</li> <li>- Publishes key figures related to the pandemic on a dashboard</li> </ul>	None recorded (under broader definition: reusers, data journalists)	<ul style="list-style-type: none"> <li>- Technical capital</li> <li>- Cultural capital (institutional prestige; reach; ability to forge partnerships with private-sector enterprises)</li> <li>- Economic capital (financial support from foundations, non-governmental organisations, university budget)</li> </ul>	Social (improved public understanding of the pandemic, democratised and easy access to relevant data may have motivated compliance with restrictions and facilitated research)

