

Data.europa.eu and Citizen-generated Data

Opportunities and challenges associated to the inclusion of citizen-generated data in data.europa.eu

data.europa.eu

The official portal for European 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

data.europa.eu

Email: info@data.europa.eu

Authors:

Oscar Corcho
Javier Jiménez Blanco
Carlos Morote
Elena Simperl

Last update: 18-04-2022
<https://data.europa.eu/>

DISCLAIMER

By the European Commission's Directorate-General for Communications Networks, Content and Technology. 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, 2022
© European Union, 2022



OA-06-22-186-EN-N

ISBN: 978-92-78-42937-9

doi: 10.2830/137589



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.

Contents

1. Introduction and motivation.....	6
2. Glossary and relevant definitions.....	8
3. Methodology.....	9
3.1. Step 1: selection of literature sources.....	9
3.2. Step 2: analysis framework for CGD on open data portals.....	10
3.3. Step 3: field analysis of CGD available in open data portals.....	10
4. Analysis framework for citizen-generated data.....	11
4.1. Proposed framework and dimensions.....	12
5. Analysis of citizen-generated data presence in several open data portals: selection process and results.....	15
5.1. Results presentation and discussion.....	16
6. Conclusions, recommendations and future work.....	19
References.....	21
A. Reviewed open data portals.....	24
A.1. National: Czechia (data.gov.cz).....	24
A.2. National: France (data.gouv.fr).....	25
A.3. National: Poland.....	25
A.4. National: Spain (datos.gob.es).....	26
A.5. National: United Kingdom (data.gov.uk).....	26
A.6. Regional: Catalonia (governobert.gencat.cat).....	26
A.7. Regional: Trent (dati.trentino.it).....	27
A.8. Local: City of Barcelona (opendata-ajuntament.barcelona.cat).....	27
A.9. Local: City of Berlin (daten.belin.de).....	28
A.10. Local: City of Dublin (data.smartdublin.ie).....	28
A.11. Local: City of Helsinki (hri.fi).....	28
A.12. Local: City of Lyon (data.grandlyon.com).....	29
A.13. Local: City of London (data.london.gov.uk).....	30
A.14. Local: City of Madrid (datos.madrid.es).....	30
A.15. Local: City of Milan (dati.comune.milano.it).....	31
A.16. Local: City of Zaragoza (zaragoza.es/sede/portal/datos-abiertos).....	31

Executive summary

Since the launch more than a decade ago of the initial set of open data portals, an increasing number of public administrations have also proposed their open data strategies and launched their open data portals to facilitate finding and using the data produced, stored and maintained by public administrations as part of their daily activities. Typical categorisations for datasets available in these data portals are those that group datasets according to their domains (agriculture, economy, education, culture, energy, transport, etc.), or that make it possible to filter them according to their formats, the quality of their metadata, their geographical or temporal coverage, their licenses, etc.

This report explores a category of data that may be considered as transversal to the aforementioned categorisations: citizen-generated data (CGD). Indeed, citizens living in or visiting Europe generate large amounts of data as part of their daily activities. Data generation and capture can be done as a result of a conscious or explicit choice by citizens, such as when they contribute to public resources like [Wikidata](#) or [OpenStreetMap](#), when they submit a claim or suggestion to their municipality website, or when they contribute or curate data for citizen science initiatives (e.g. recording temperature values or air quality measurements in their neighbourhood and validating or curating data about trees in a city). Data capture can be unconscious or implicit, that is, without necessarily realising that such data is being generated and stored, despite having accepted general data protection regulation-related conditions or cookies in websites. For example, this happens when the geolocation of citizens is recorded as they move around with their mobile phones or when they scan their transport card when entering into public transport.

Much of this CGD is managed by private companies. Citizens provide their consent to have such data captured, since they normally benefit from free-to-use services offered by these organisations. In turn, these organisations gain the right to use such data to improve their services or provide other added-value services or data products based on such data (e.g. business ratings and occupancy data in Google Places and FourSquare, FixMyStreet reports on graffiti and potholes, etc.). CGD may be also managed by non-profit foundations and associations, with a public good objective (e.g. Wikipedia, Wikidata, OpenStreetMap). And ultimately, they may be managed by public administrations.

However, these types of data sources are not widely represented in open data portals (at any administrative level, from municipalities to data.europa.eu), with only some exceptions (Balestrini, Mara and Kotsev, 2021). The main reason for this may be that, by design, open data portals have traditionally limited themselves to publishing data directly generated and managed by public administrations. Therefore, it is a good moment to reinvestigate the opportunities and challenges that including this type of CGD in open data portals may bring in.

The main objective of this report is to provide some insight about CGD that may be part of open data portals, allowing public administrations to better understand whether and how CGD may be made available as part of their open data offering. The report provides a categorisation of different types and dimensions of CGD that may be offered in open data portals (and that is already offered in some of them), together with a discussion on the opportunities and challenges that their inclusion in data.europa.eu may bring in.

The main take aways from this study are the following.

- In the context of open data portals in Europe some CGD datasets are available, but the data generated by citizen science (CS) projects are rarely included.
- Many open data portals include datasets that are generated as a result of citizen participation initiatives (surveys, complaints, etc.). Even though these datasets contain CGD, these initiatives are typically initiated by administrations and citizens are relegated to mere providers.
- There are interesting opportunities to include CGD in open data portals, in general, and in data.europa.eu in particular. Following these recommendations, the amount of CGD would not only increase but would also potentially promote citizen participation.

Note on data availability

Data produced during the development of this study, which serves as the base for the statistical analysis presented in Section 5, is available as open data [here](#). This also includes a register of the selected papers during our desk research. If the reader is interested in some studied datasets, a complete list of them is provided in Annex A.

1. Introduction and motivation

Open government portals in Europe and beyond have been set up to facilitate access to and the reuse of public sector information (PSI). PSI is subject to the EU's open data directive, where it is defined as 'any kind of information that is produced and/or collected by a public body and is part of the institution's mandated role' (European Parliament and Council, 2003, 2013, 2019).

According to this definition, the release of PSI datasets is driven by public authorities and linked to their remit and operations. This understanding of public sector data, alongside the practicalities of getting the first data portals off the ground, has had an impact on the types of datasets that can be found today in most open government data portals at European, national and regional levels, and on data.europa.eu. Portals took different approaches to choosing and curating the data they publish: some are rather prescriptive with respect to whom they allow to put data on the portals, others, for instance the French national open governmental data portal ⁽¹⁾, take a more inclusive stand by allowing citizens to upload their own datasets, and versions of those that are already available. As they advanced with the implementation of the open data directive, portals increasingly recognised the need to build data communities, bringing together publishers and users and blending the differences between the two, to facilitate data use and impact, understand the needs of the data ecosystem and spread data maintenance costs (Ibanez et al., 2020).

This trend of opening up publishing processes to ensure impact and sustainability is well aligned with other data policies and regulations in the EU, particularly in the area of data sharing. These policies and regulations are meant to empower various stakeholders (government, businesses, citizens) to give access to and use non-personal data (some open, some not) to create value and make better decisions. Some developments in this space include the 2018 appointment of the expert group on business-to-government data sharing, whose final report (European Commission, 2021a) was published in February 2021, and the European data strategy, with components around data governance, common data spaces and, as of February 2022, the European data act. The aim is 'to make more data available for use, and set up rules on who can use and access what data for which purposes across all economic sectors in the EU. The new rules are expected to create €270 billion of additional GDP for EU Member States by 2028 by addressing the legal, economic and technical issues that lead to data being underused' ⁽²⁾.

Against this background, this analytical report analyses the opportunities and challenges associated with the inclusion of citizen-generated data (CGD) in established and emerging data publishing and sharing frameworks. This includes a range of data scenarios, which current frameworks overlook, from participatory sensing to crowdsourced geospatial datasets. The technologies, processes and practices used in CGD initiatives are yet to become fully integrated with open data portals and, in the future, common data spaces. This is a lost opportunity in more than one way: existing citizen-generated datasets are less impactful, while portals and spaces miss out on a critical stakeholder group.

⁽¹⁾ <https://www.data.gouv.fr/en/>

⁽²⁾ https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en, accessed in March 2022.

Our analysis focuses on CGD that does not contain personal data or where personal data is conveniently anonymised or aggregated. That is, such data is orthogonal to the application of the general data protection regulation across the EU (European Parliament and European Council, 2016) and the data protection laws and norms across Member States. Some proposed technical solutions would allow individuals to manage their own data pods (e.g. SOLID data pods, as described in <https://solidproject.org/>), but the role of established open data portals in publishing or publicising such datasets remains underexplored. Furthermore, it is unclear how existing data licensing and governance models popular with current open government datasets held by public authorities co-exist with citizen-centric approaches such as data commons (Heath et al., 2021) and participatory data stewardship (Patel, 2021).

When discussing CGD, we distinguish between ‘explicit’ and ‘implicit’ data generation. In the former, citizens engage with various organisations and initiatives, including, but not restricted to, public authorities, to collect and curate data. Examples include contributions to geospatial datasets such as OpenStreetMap and Wikidata, or to CS projects such as Restart, where the public submit data about repairing different types of devices via an app to reduce waste ⁽³⁾. In the latter, citizens do not actively submit data – rather, data is generated by activities they undertake when using public services and smart or mobile devices and sensors. Examples include aggregated datasets of public-transport journeys, footfall data and various types of environmental datasets created through participatory sensing.

No matter how they are created, citizen-generated datasets are routinely used by public authorities and private and third sector organisations to complement the data they already have. They can offer much better coverage and quality at lower costs compared to traditional, less decentralised data collection methodologies. How this works was showcased, for instance, in H2020 Qrowd, where the City of Trent, Italy, used a mix of volunteers equipped with location-based services and online crowdsourcing to create eight high-value mobility datasets at 25 % of the costs of alternative methods (Maddalena et al., 2020). The opportunities are increasingly being recognised in policy work as well: a recent report by the H2020 ACTION project (Notermans et al., 2022) discusses the recommendations of six policy masterclasses from the EU, Italy, the Netherlands, Norway, Spain and the United Kingdom. Among others, it argues for ‘the development of an open data platform between policymakers and citizen science projects to allow for aligned, common and transdisciplinary data structures’. This report also discusses, in the specific case of Norway, the potential use of the Norwegian national open data platform, data.norge.no, as a means to integrate, sort and analyse CS data flows, and exchange best practices on existing tools and data management protocols.

Outside the scope of this report will be the analysis of data people produce when using digital services owned by businesses (e.g. Foursquare, Waze and Citymapper) that hold private crowdsourced datasets and use them in accordance with users’ consent.

This report is structured as follows. In Section 2 we compile terms and definitions that are commonly used in the context of CGD, together with some relevant references. In Section 3 we describe the

⁽³⁾ <https://therestartproject.org/data-and-insights/>

methodology we followed to understand ongoing opportunities and challenges around CGD. This resulted in a framework to classify and characterise citizen-generated datasets (Section 4), which we applied on a selection of open government data portals to get a sense of how well they cover such datasets (Section A). We conclude with core recommendations and plans for future work in Section 6, which includes validating the framework with portal providers and expanding the scope of the portal analysis.

2. Glossary and relevant definitions

In this section, we compile several relevant terms and definitions that are used throughout the report, with references to relevant literature sources where such definitions have been coined or are further explained.

Citizen-generated data (CGD). There is no single widely-agreed definition for CGD in the literature. For instance, CGD is defined as ‘the data that individuals consciously generate and that are openly available for use in the public domain’ by Meijer and Potjer (2018).

Sieber and Johnson (2015) considers CGD as an important form of open data for public governance, as it facilitates collaborative action by various stakeholders to tackle public concerns. Similarly, a recent Joint Research Centre (JRC) report on CGD for public policy (Ponti and Craglia, 2020) subscribes to a notion of CGD as ‘data that people or their organisations produce to directly monitor, demand or drive change on issues that affect them’ (DataShift, 2015). Further on, DataShift (2015) notes that CGD ‘is actively given by citizens, providing direct representations of their perspectives and an alternative to dataset collected by governments or international institutions’. Finally, another definition, contextualised around climate change and with a wider scope beyond more traditional data structures, is given by Suman et al. (2020): citizen-generated datasets are ‘rhetorical resources actively produced or gathered by citizens for making an argument to support environmental justice claims. [CGD enables lay citizens to] question expert knowledge production through critical making tactics and creates opportunities to generate credible public science’. Following this line of thought, the authors suggest that CGD could be regarded as a form of social protest that challenges the institutional status quo and politics and politicisation of data (Suman et al., 2020).

As can be seen, the definition of CGD is highly dependent on the context. Therefore, for the purposes of this work, the broad definition of Meijer and Potjer (2018) is employed, although we will expand it to data that is also provided unconsciously by citizens, while these other factors that the rest of the definitions point out as relevant are addressed through the use of dimensions as ‘primary/secondary’ or ‘by/about’ (see Section 4.1).

Citizen science (CS). According to the European Citizen Science Association (ECSA), this is an umbrella term that describes a variety of ways in which the public participates in science, with the following characteristics: citizens are actively involved in research, in partnership or collaboration with scientists or professionals; and there is a genuine outcome, such as new scientific knowledge, conservation action or policy change.

Volunteer contributions most commonly consist of data observations, or analyses, which are used by scientists and policymakers (Kullenberg and Kasperowski, 2016). There are a series of terms with similar meanings, including ‘community-based monitoring’ (Danielsen et al., 2005), ‘volunteer

monitoring' (Shirk et al., 2012), 'participatory science' (Ashcroft et al., 2012) and 'crowdsourced science' (Watson and Floridi, 2018), all designating the contribution of non-scientists to different scientific disciplines.

Crowdsourcing is a high-level concept that refers to a decentralised approach to problem solving that seeks ideas and contributions from a typically open, large and evolving group of participants. In CGD, the problem to be solved is the collection of datasets with certain parameters (e.g. coverage, number of observations and level of quality) or the enrichment of an existing dataset (e.g. adding labels or classifications to scientific datasets). As noted earlier, CS is sometimes referred to as crowdsourced science. Such projects represent a type of distributed scientific collaboration, also known as collaboratories, but until recently most of this research has focused on projects in which all contributors are scientists or supporting professionals. Emerging scientific cyberinfrastructure projects are a related phenomenon that bear a stronger similarity to CS than do the collaboratories discussed in the literature to date (Wiggins and Crowston, 2011). Beyond science, many datasets in disaster response, mapping and healthcare are created via crowdsourcing – one example is the Zoe app, the world's largest database of COVID-19 symptoms, collected from citizens⁽⁴⁾. Another example is OpenStreetMap, where volunteers create alternatives to commercial products such as Google Maps. There are many other terms similar to crowdsourcing, including 'collective intelligence' (Leimeister, 2010), 'wisdom of the crowds' (Kittur et al., 2007) and 'peer-production systems' (Benkler, 2016).

Open science 'is a framework for how scientists interact with one another and how the public engages with, and is engaged in, science' (Hecker et al., 2018). The European Open Science Cloud is the most relevant ongoing initiative in this context for Europe. Data is a key pillar of any open-science strategy, alongside methods and research management tools. This includes open access, which is driven by the understanding that publicly funded research should be accessible to all members of society. The open science imperative of sharing information and results from publicly funded research has led to the promotion of the open access publication model (where scientific publications are freely available rather than subject to expensive subscription rates) as well as open scientific data repositories (where datasets are made freely available to other potential users). Some CS datasets are published in such repositories.

3. Methodology

We followed a methodology consisting of three steps: selection of literature sources; design of an analysis framework; and field analysis of open government data portals using the framework.

3.1. Step 1: selection of literature sources

First, we have made a systematic literature review by considering the following sources: [Web of Science](#)⁽⁵⁾, as the most commonly used source for documents in the scientific domain, and the [JRC Publications Repository](#)⁽⁶⁾, given the context of this study and the fact that some relevant documents may be available here but not in the Web of Science. The goal was to identify relevant papers discussing

⁽⁴⁾ <https://covid.joinzoe.com/>

⁽⁵⁾ Website that offers access to different databases of reports and citations; owned by Clarivate.

⁽⁶⁾ The JRC is a Directorate-General of the European Commission, and its repository provides public access to all state-of-the-art JRC science publications.

the generation, publication and use of CGD in open data portals. The literature review was also useful to compile the definitions from Section 2.

We searched literature (published papers, technical reports, white papers and official publications) with the following topics: ‘citizen-generated data’, ‘citizen science’, ‘open science’ and ‘open data’. We considered the title, abstract and keywords (both author-generated and automatically generated) to establish topic relevance. For reproducibility purposes, the query used on Web of Science is provided as a footnote (?).

Through this search we found 146 Web of Science papers and 19 JRC Science Hub ones. We reviewed first the title and abstract, using the following inclusion criteria: the document should contain clear references to the (potential) publication of CGD as open data; and the document should discuss potential categories or dimensions to classify or characterise CGD. Where the title and abstract were inconclusive, we also read the introduction and conclusion sections.

The analysis led to the following set of papers:

- Meijer and Potjer, 2018
- Suman et al., 2020
- Kullenberg and Kasperowski, 2016
- Manzoni et al., 2021
- Attard et al., 2015
- Chatfield and Reddick, 2017
- Ponti and Craglia, 2020.

From these papers, we considered their references (using snowballing) in order to identify additional papers that were not on the list above. We found one paper which was not available on Web of Science and the JRC Science Hub (DataShift, 2016).

3.2. Step 2: analysis framework for CGD on open data portals

We reviewed the list of papers from Step 1 in detail to understand how they refer to, describe or characterise CGD. For each paper we compiled dimensions for our analysis framework, following a similar approach to European Commission (2021b). As commented previously, this approach is useful to deal with the multiplicity of aspects that are involved in the constantly evolving concept that is CGD. The results are discussed in Section 2, where a characterisation of CGD can be found, and in Section 4, where these dimensions are studied.

3.3. Step 3: field analysis of CGD available in open data portals

We selected a set of European open data portals through purposive sampling. The sample includes portals at different administrative levels (local, regional and national) and from different parts of Europe.

(?) <https://www.webofscience.com/wos/allodb/summary/7a543993-d5ad-473e-8066-b491954083c8-250f59dc/relevance/1>

The sample was bootstrapped with the papers from Step 2, which mentioned some open data portals as publishers of citizen data. We extended the sample by consulting CS experts via organisations such as ECSA.

Our aim was not to rank portals based on their support for CGD; for this initial study, we first wanted to analyse the status quo qualitatively using the framework from Step 2 as a guide. A larger sample, including quantitative analyses to understand specific details of CGD publishing in the European data ecosystem, could be an interesting extension as CGD becomes more widespread in open government data.

We have used keywords such as ‘citizen contribution’, ‘citizen participation’, ‘citizen science’, ‘citizen generated data’ and ‘complaints-suggestions’, in English or in their local languages. We then checked the datasets returned by these queries to see if they could be considered CGD and applied the dimensions from our analysis framework from Step 2.

4. Analysis framework for citizen-generated data

In the first step of our methodology we selected papers relevant to the CGD domain. In the second step, we reviewed the papers in detail to derive our framework of analysis. Two authors of the report read the papers independently to identify dimensions of analysis and compared the results, which were in large agreement. Then the entire team discussed their findings and agreed on a shared terminology across the papers, which we use to analyse the portals later on.

We can distinguish among three types of studies: those where CGD is considered yet another type of open government data, and hence share the same characteristics; those focused on the potential use of CGD for policymaking; and those looking at CGD as a by-product of CS.

Considering CGD as **open government data**, Attard et al. (2015) discuss characteristics related to the data (nature and accountability) and to the data portal (transparency, accessibility and openness), along with legal obligations, institutional arrangements and public engagement (participation and collaboration). The latter is the most interesting for our analysis, as it hints at building data communities and ecosystems where a range of stakeholders are involved in the entire data lifecycle. A similar approach is presented in Chatfield and Reddick (2017).

For **CGD for policymaking**, Suman et al. (2020) reports the following dimensions.

- **Initiator(s)**. The people or organisations that initiate the data collection process.
- **Policy uptake**. Whether CGD has had any impact in policy uptake (yes/no), and if yes, the intended policy use(s) and the actual policy uptake and use.
- **Legitimation strategies**. Whether the CGD process identifies any legitimisation strategy for the data that is generated.

An explorative comparative case analysis of 25 cases of citizen-generated open data is presented in Meijer and Potjer (2018) with the objective of providing an initial understanding of the variation of

practices, the roles of different actors and the potential contribution of these practices to the public domain. The main dimensions explored are related to:

- the conditions under which citizens are willing to provide data;
- the people or organisations that hold and manage the data;
- the impact of the open-data initiative on government practices.

In the context of **scientific research and CS**, the secondary study presented in Kullenberg and Kasperowski (2016) analyses publications that discuss the use of CGD in different research areas. The main objective is to identify common themes in CS initiatives and the scientific datasets they produce. The work concludes that in areas such as biology, conservation and ecology, CS initiatives act mainly as a method to collect and classify data. In areas such as geographic information research, citizens participate mostly in the collection of geographic data, but not in its classification. In the context of social sciences and epidemiology, CS is a tool that facilitates studying and public participation in areas such as environmental issues and health. Complementarily, Manzoni et al. (2021) reports the results of an online survey with 45 respondents on CS approaches and initiatives in Europe. The survey does not focus on data specifically, but discusses each CS initiative in terms of different dimensions; being *Involved actors and their roles and scope of intervention, Tools and methodologies, Areas of application* and *Impact* those of more relevance when studied in the context of Open data portals.

Finally, Ponti and Craglia (2020) analyses different CS initiatives where CGD is being generated, with an additional focus on CGD for policymaking, according to the following dimensions:

- purpose of data collection;
- primary categories of the project;
- type of data collected and technology used for collection;
- output of data collection;
- data owner;
- whether the CGD is being used by the public sector;
- other criteria related to the FAIR – Findable, Accessible, Interoperable, Reusable – data principles (Wilkinson et al., 2016).

The paper surveys 18 projects, concluding that in most cases the generated data is not used directly by the public sector, nor published by them.

4.1. Proposed framework and dimensions

Following the review of the papers, we propose the following framework of analysis to understand challenges and opportunities for CGD in open government data portals.

ID	Dimension	Dimension values	Description
P1	Percentage of CGD datasets in the portal	Percentage	This dimension measures the percentage of datasets published in the open data portal that can be categorised as CGD
P2	Existence of general guidelines for publishing CGD	Yes/No (y/n)	This dimension considers whether the open data portal provides any guidelines for any of the involved

			actors (public administration, citizens, reusers, etc.) on how to publish or make use of CGD datasets
P3	Availability of methods and tools for CGD management and use	Yes/No (y/n)	This dimension evaluates whether there are any methods or tools available in the portal (or referenced by it) for the acquisition, curation or exploitation of CGD. This may also include web and mobile apps to ease citizen participation, including data contributions and data visualisations. The following codes are used: m: maps; qe: query engines; ud: upload datasets (ability for citizens to upload datasets); cd: contribute to datasets (ability for citizens to contribute to datasets); api: APIs to contribute data)
P4	Proposal of quality assurance mechanisms for CGD	Yes/No (y/n)	This dimension focuses on whether there are any specific procedures to control the quality of CGD
D1	Area	gl: geolocalisation or geospatial data e: environmental data cs: citizen science data v: votes su: surveys qa: questions/answers (it includes suggestions and complaints) st: statistics	This dimension identifies the general domain in which the dataset can be categorised. These areas have been selected as they are the most typical areas where CGD can be found in open data portals, as obtained from our literature review and discussions with experts
D2	Actor roles	c: citizens pa: public administration o: other	This dimension identifies which type of actor is involved in different roles related to the dataset. This will include an analysis of who are the initiators (citizen initiatives, NGOs, research centres, governments, etc.), the funders (research funding organisations, public administrations, foundations, etc.), the data providers (commonly, citizens), the curators and the keepers
D3	Specific CGD dataset guidelines	Yes/No (y/n)	This dimension analysis of whether the dataset has any reference to specific guidelines on how citizens can contribute to it
D4	Primary/secondary CGD	p/s	This dimension identifies whether the CGD dataset has been created as the primary purpose, or whether it is a secondary result of some other activity
D5	Data by or about citizens	by/about	This dimension analyses whether citizens are consciously providing their data (e.g. curating a dataset about trees in their neighbourhood), or non-consciously by making use of some service or

			providing the data for another purpose (e.g. using their transport cards to access public transport)
D6	Expected policy or operational impact	Yes/No (y/n)	This dimension considers whether the dataset has been generated or is being curated with a specific policy or operational impact in mind
D7	Data format	XML, JSON, CSV, XLSX, ZIP or others	This dimension analyses the format in which the data is made available
D8	Licensing	Open/Closed (o/c)	This dimension focuses on the type of license associated to the dataset. In the vast majority of cases, it is expected that these will be open licenses, since we are considering datasets published in open data portals

A brief description of all of them is provided below, to clarify and delimitate their meanings and areas of applicability.

- **Open data portals dimensions.** These are the dimensions that are applicable to open data portals as a whole, in relation to their treatment of CGD.
 1. **Percentage of CGD datasets in the portal.** This dimension measures the percentage of datasets published in the open data portal that can be categorised as CGD.
 2. **Existence of general guidelines for publishing CGD.** This dimension considers whether the open data portal provides any guidelines for any of the involved actors (public administration, citizens, reusers, etc.) on how to publish or make use of CGD datasets.
 3. **Availability of methods and tools for CGD management and use.** This dimension evaluates whether there are any methods or tools available in the portal (or referenced by it) for the acquisition, curation or exploitation of CGD. This may also include web and mobile apps to ease citizen participation, including data contributions and data visualisations. The following codes are used: m: maps; qe: query engines; ud: upload datasets (ability for citizens to upload datasets); cd: contribute to datasets (ability for citizens to contribute to datasets); api: APIs to contribute data).
 4. **Proposal of quality assurance mechanisms for CGD.** This dimension focuses on whether there are any specific procedures to control the quality of CGD.
- **CGD dataset dimensions.** These are the dimensions that are applicable and will be used to evaluate specific datasets that can be considered as CGD. For each of them in a portal, we will evaluate:
 1. **Area.** This dimension identifies the general domain in which the dataset can be categorised. The following codes will be used: gl: geolocalisation or geospatial data; e: environmental data; cs: CS data; v: votes; su: surveys; qa: questions\answers, including suggestions and complaints; st: statistics. These areas have been selected as they are the most typical areas where CGD can be found in open data portals, as obtained from our literature review and discussions with experts.
 2. **Actor roles.** This dimension identifies which type of actor is involved in different roles related to the dataset, using the following code scheme: c: citizens; pa: public administration; o: other. This will include an analysis of who are the 'initiators' (citizen

initiatives, NGOs, research centres, governments, etc.), the ‘funders’ (research funding organisations, public administrations, foundations, etc.), the data providers (commonly, citizens), the ‘curators’ and the ‘keepers’.

3. **Specific CGD dataset guidelines.** This dimension analyses whether the dataset has any reference to specific guidelines on how citizens can contribute to it.
4. **Primary or secondary CGD.** This dimension identifies whether the CGD dataset has been created as the primary purpose, or whether it is a secondary result of some other activity.
5. **Data by or about citizens.** This dimension analyses whether citizens are consciously providing their data (e.g. curating a dataset about trees in their neighbourhood), or non-consciously by making use of some service or providing the data for another purpose (e.g. using their transport cards to access public transport).
6. **Expected policy/operational impact.** This dimension considers whether the dataset has been generated or is being curated with a specific policy or operational impact in mind.
7. **Data format.** This dimension analyses the format in which the data is made available.
8. **Licensing.** This dimension focuses on the type of license associated to the dataset. In the vast majority of cases, it is expected that these will be open licenses, since we are considering datasets published in open data portals. We will use the following codes: o: open; c: closed.

5. Analysis of citizen-generated data presence in several open data portals: selection process and results

In this section, we identify first the set of open data portals that have been explored as part of our field work, where we have applied our framework of dimensions globally and for specific CGD datasets that have been identified, and describe the main results and conclusions that can be derived from this analysis.

We considered the following portals, which are a mix of portals mentioned in the papers reviewed and additional ones, as explained in the methodology:

National portals:

1. Czechia (data.gov.cz)
2. France (data.gouv.fr)
3. Poland (gov.pl)
4. Spain (datos.gob.es)
5. United Kingdom (data.gov.uk)

Regional portals:

1. Catalonia, Spain (transparenciacatalunya.cat)
2. Trent, Italy (dati.trentino.it)

Local portals:

1. Barcelona, Spain (opendata-ajuntament.barcelona.cat)
2. Berlin, Germany (daten.berlin.de)

3. Dublin, Ireland (data.smartdublin.ie)
4. Helsinki, Finland (hri.fi)
5. London, UK (data.london.gov.uk)
6. Lyon, France (data.grandlyon.com)
7. Madrid, Spain (datos.madrid.es)
8. Milan, Italy (dati.comune.milano.it)
9. Zaragoza, Spain (zaragoza.es)

Annex A provides details about these open data portals, the keywords and tags, when available, that have been used to identify CGD datasets in them, and the final set of datasets that have been considered in our analysis according to our framework of dimensions.

5.1. Results presentation and discussion

The percentage of CGD available in the explored open data portals is rather low in comparison with the rest of available datasets (Figure 5.1). This is not a surprise, as it was expected from our initial literature survey and from our informal discussions with people responsible for open data portals. Therefore, the average percentage is below 1 % (0.83 %). Some portals, e.g. Czechia or the United Kingdom, are even lower (0.02 %). It is worth mentioning the case of Trent, where the high percentage of CGD is only due to the reuse of GeoNames in order to provide geolocalised data; and the one of France, that even when it includes 17 datasets identified as CGD, the percentage shown in the graph is below 0.5 % due to the total amount of datasets that this portal gathers (over 40 000 datasets).

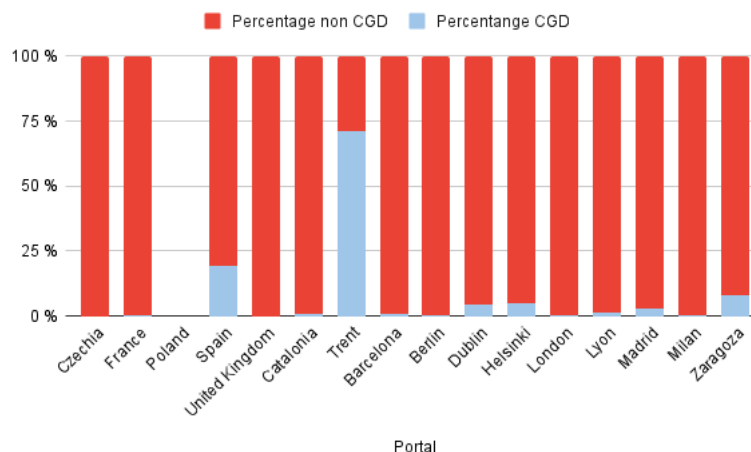


Figure 5.1: Percentage of CGD in each studied portal in relation to all datasets that are published

Most of the initiators (and funders) are ‘public administrations’, as shown in Figure 5.2. They also act as curators and keepers in virtually all of the cases. An exception to this situation is the Spanish portals, where projects initiated by citizens are less uncommon.

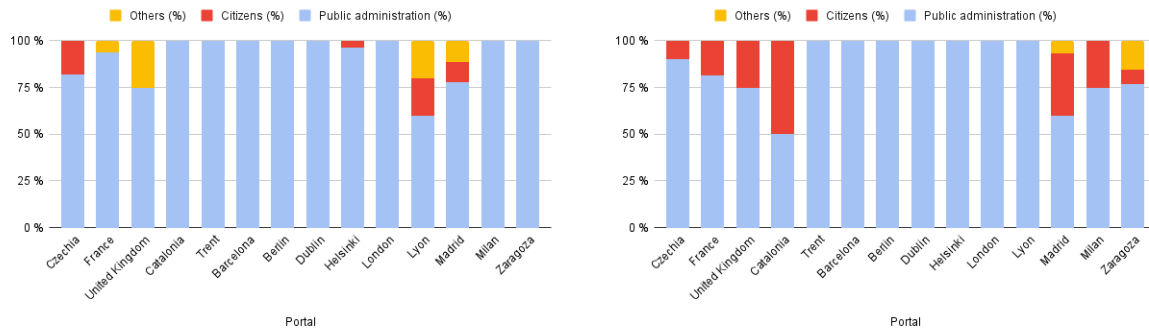


Figure 5.2: Proportion of actors involved in CGD data that are published in open data portals. Proportion of project funders by portal (left). Proportion of project initiators by portal (right).

The most frequent area where CGD is produced is ‘questions and answers’. Due to the design of these dimensions, this also includes participatory projects such as participatory budgets. Other areas include ‘surveys’ and ‘statistics’. This trend is even clearer in portals with large amounts of CGD, such as the ones of Zaragoza or Helsinki. In addition, ‘geolocalisation’ is also relevant because user-generated-content projects, such as GeoNames or OpenStreetMap, are usually reutilised to provide geolocalised data. Figure 5.3 summarises these conclusions.

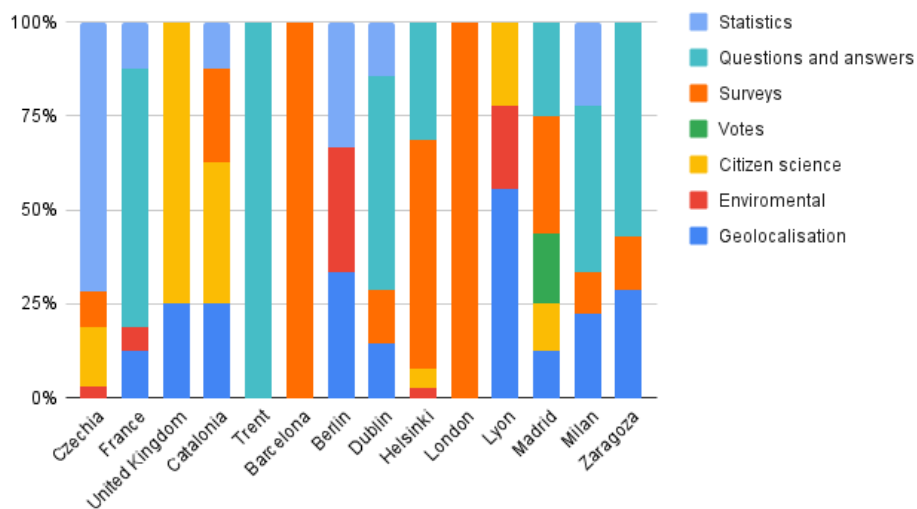


Figure 5.3: CGD datasets per Portal and Area

Unexpectedly, primary CGD is more frequent than secondary; and there are more data created by citizens than merely about them. A relevant case in this aspect is the French portal, where it is possible to find many participatory projects that involve CGD. This serves as an example for other open data portals. Figure 5.4 intends to provide useful information in this context.

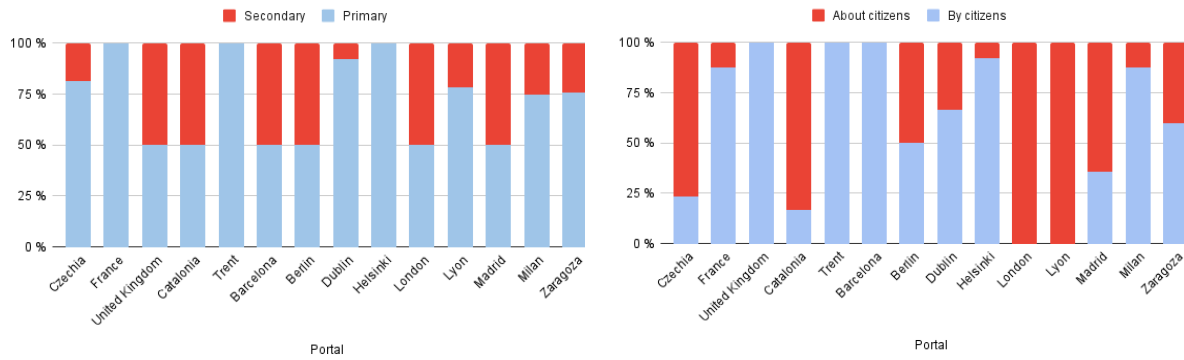


Figure 5.4: Citizen’s role in CGD projects in open data portals in Europe. Primary/secondary CGD (left). By/about CGD (right)

Regarding formats, licensing and accessibility, almost all portals include only data available with open licenses, with the exception of the UK. In addition, the amount of datasets provided in linked data formats (e.g. JSON, XML and RDF) is quite high (Figure 5.5).

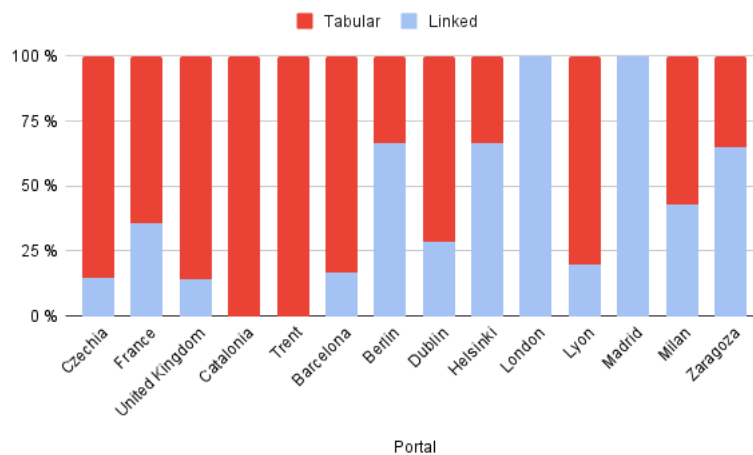


Figure 5.5: Formats of datasets by portal

Finally, it is worth commenting that almost all portals include some general guidelines on how to access the catalogue and how to use the provided tools – the Polish portal seems to be an exception – but none of them include specific references on how to access or operate with the CGD that they contain. Only a few of the portals allow citizens to contribute their datasets, such as the French national portal, with options for citizens to upload their own versions of the datasets, or the portal of the Province of Trent in Italy, where only local administrations are allowed to contribute. However, none of them offer options to modify collaborative datasets online. Finally, none of the portals include documentation about how to ensure the quality of CGD, nor mechanisms to control it.

6. Conclusions, recommendations and future work

Based on this initial analysis, we derived the following conclusions.

- C1.** All portals had very few citizen-generated datasets, both in absolute terms and relative to the number of datasets hosted by the portals.
- C2.** In most cases, citizens are primarily involved in generating or collecting the data, but the remaining work required to publish the data is driven by public administrations. The efforts are initiated by public administrations rather than bottom-up by citizens, who are also less involved in curating or maintaining the data.
- C3.** The most frequent areas of CGD published in open data portals are: 'questions and answers', 'surveys' and 'statistics'. This is complementary to CS datasets (Ponti and Craglia, 2020).
- C4.** Primary CGD is more common than secondary CGD in open data portals.
- C5.** Most of CGD that is made available is shared with open licenses.
- C6.** Almost 30 % of the studied CGD datasets are available in open formats like JSON and XML. A much smaller percentage uses proprietary formats, typically XLSX.
- C7.** None of the studied portals included documentation about how to contribute and use CGD, nor about specific procedures to ensure data quality in this context. In fact, CGD is not explicitly identified as a data collection approach.
- C8.** In general terms – some exceptions are discussed in Section 5 – portals do not offer tools to facilitate citizen contributions, either at the level of datasets (upload their own data) or individual records (change, curate and maintain existing data).
- C9.** Following from **C2** and **C8**, we could find no evidence of participatory approaches to design data pipelines or collect and implement feedback from citizens on broader data strategy.
- C10.** No general guidelines on how to govern CGD in open data portals are provided; this seems to limit the emergence of more of these types of datasets.

Based on our analysis, we provide the following recommendations for primary open government data sources and intermediaries such as data.europa.eu. The aim is to increase and improve the presence of CGD on open data publishing beyond open science, but also involve citizens in designing open data policy, processes and governance.

- R1.** Actively seek valuable CGD assets through open calls and partnerships with key CS players such as ECSA, and national and regional offices in CS, and CS projects.
- R2.** Facilitate the discovery of CGD in open data portals by tagging all CGD datasets with a specific tag such as 'CGD' or 'citizen-generated data' (e.g. Portal of Dublin, although the tagging should be made more specific).
- R3.** Include keywords/tags in official languages of the EU to facilitate comparative studies using multiple datasets.
- R4.** Establish procedures to capture CGD processes and data validation methods to increase the trust of third-party data users.
- R5.** Extend data and metadata quality capabilities with metrics specific to CGD (Hunter et al., 2013).

- R6.** Include CGD aspects in upcoming open data reports, such as the European analysis of open data portals (European Commission, 2021b).
- R7.** Collect new and tag existing use cases from data.europa.eu to showcase the value of CGD datasets (e.g. the French national portal).
- R8.** Link use cases to applications and co-locate tools and documentation to encourage reuse by diverse audiences, including people with varying levels of data literacy.
- R9.** Create tools and applications that consume this type of data and allow citizens to contribute – via data collection or curation – to the original data sources.
- R10.** Allow citizens to contribute information within the portal, not only allowing the upload of complete datasets, but also the addition or maintenance of instances to existing records.

This report is the first in a line of research concerned with sources and types of open data that are currently, for historical or other reasons, overlooked in open governmental data portals and data.europa.eu. In future studies, we will carry out interviews with open government data publishers, along with more established CGD stakeholders (e.g. citizens, scientists and civil society), to complement the current analysis, identify synergies and put forward proposals for specific projects to add more, high-quality CGD to existing portals, and assess its use and impact.

References

- Ashcroft, M. B., Gollan, J. R. and Batley, M. (2012), 'Combining citizen science, bioclimatic envelope models and observed habitat preferences to determine the distribution of an inconspicuous, recently detected introduced bee (*Halictus smaragdulus* Vachal Hymenoptera: Halictidae) in Australia', *Biological Invasions*, Vol. 14, No 3, Springer Nature, pp. 515–527. <https://doi.org/10.1007/s10530-011-0092-x>
- Attard, J., Orlandi, F., Scerri, S. and Auer, S. (2015), 'A systematic review of open government data initiatives', *Government Information Quarterly*, Vol. 32, No 4, Elsevier, pp. 399–418. <https://doi.org/10.1016/j.giq.2015.07.006>
- Balestrini, M., Kotsev, A., Ponti, M. and Schade, S. (2021), 'Collaboration matters: capacity building, up-scaling, spreading, and sustainability in citizen-generated data projects', *Humanities and Social Sciences Communications*, Vol. 8, Article No 169, Springer Nature. <https://doi.org/10.1057/s41599-021-00851-5>
- Benkler, Y. (2016), 'Peer production and cooperation', *Handbook on the Economics of the Internet*. Edward Elgar Publishing, pp. 91–119. <https://doi.org/10.4337/9780857939852.00012>
- Chatfield, A. T. and Reddick, C. G. (2017), 'A longitudinal cross-sector analysis of open data portal service capability: the case of Australian local governments', *Government Information Quarterly*, Vol. 34, No 2, Elsevier, pp. 231–243. <https://doi.org/10.1016/j.giq.2017.02.004>
- Danielsen, F., Burgess, N. D. and Balmford, A. (2005), 'Monitoring Matters: examining the potential of locally-based approaches', *Biodiversity and Conservation*, Vol. 14, No 11, Springer Nature, pp. 2507–2542. <https://doi.org/10.1007/s10531-005-8375-0>
- DataShift (2015), *What is citizen-generated data and what is the DataShift doing to promote it?*, CIVICUS. http://civicus.org/images/ER%20cgd_brief.pdf
- DataShift (2016), *Citizen-generated data and governments. Towards a collaborative model*, CIVICUS. <http://civicus.org/images/citizen-generated%20data%20and%20governments.pdf>
- European Commission, Directorate-General for Communications Networks, Content and Technology (2021a), *Towards a European strategy on business-to-government data sharing for the public interest: final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing*, Publications Office of the European Union. <https://doi.org/10.2759/406717>
- European Commission, Directorate-General for Communications Networks, Content and Technology (2021b), *Open Data Maturity Report 2021*, Publications Office of the European Union. https://data.europa.eu/sites/default/files/landscaping_insight_report_n7_2021.pdf
- European Parliament and Council (2003), Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information, *Official Journal of the European Union*, L 345, Office for Official Publications of the European Communities.

- European Parliament and Council (2013), Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information, *Official Journal of the European Union*, L 175, Publications Office of the European Union.
- European Parliament and Council (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*, L 172, Publications Office of the European Union.
- European Parliament and Council (2016), Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, L 119, Publications Office of the European Union.
- Heath, A. P. et al. (2021), 'The NCI genomic data commons', *Nature Genetics*, Vol. 53, No 3, Springer Nature, pp. 257–262. <https://doi.org/10.1038/s41588-021-00791-5>
- Hecker, S., Haklay, M., Bowser, A., Makuch, Z., Vogel, J. and Bonn, A. (2018), *Citizen Science – Innovation in open science, society and policy*, UCL Press, London, United Kingdom. <https://www.uclpress.co.uk/products/107613>
- Hunter, J., Alabri, A. and van Ingen, C. (2013), 'Assessing the quality and trustworthiness of citizen science data', *Concurrency and Computation: Practice and Experience*, Vol. 25, No 4, John Wiley & Sons Ltd., pp. 454–466. <https://doi.org/10.1002/cpe.2923>
- Ibanez, L.-D., Hoffman, M., Walker, J. and Simperl, E. (2020), *Sustainability of Open Data Portal Infrastructures: A distributed version control approach to creating portals for reuse*, Publications Office. <https://doi.org/10.2830/151838>
- Kittur, A., Chi, E., Pendleton, B. A., Suh, B. and Mytkowicz, T. (2007), 'Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie', *World Wide Web*, Vol. 1, No 2, Springer Nature. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.212.8218&rep=rep1&type=pdf>
- Kullenberg, C. and Kasperowski, D. (2016), 'What is citizen science? – a scientometric meta-analysis', *PLoS ONE*, Vol. 11, No 1, PLOS, San Francisco, California, United States. <https://doi.org/10.1371/journal.pone.0147152>
- Leimeister, J. M. (2010), 'Collective intelligence', *Business & Information Systems Engineering*, Vol. 2, No 4, Springer Nature, pp. 245–248. <https://doi.org/10.1007/s12599-010-0114-8>
- Maddalena, E., Ibáñez, L.-D. and Simperl, E. (2020), 'Mapping points of interest through street view imagery and paid crowdsourcing', *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 11, No 5, Association for Computing Machinery, pp. 1–28. <https://doi.org/10.1145/3403931>

- Manzoni, M., Vohland, K. and Schade, S. (2021), *Exploring Citizen Science Strategies and Initiatives in Europe: report on outcomes in Europe: Technical Report on Outcomes*, Joint Research Centre. https://ec.europa.eu/jrc/communities/sites/default/files/jrc123471_exploring_citizens_science_strategies_and_initiatives_jrc_for_publication.pdf
- Meijer, A. and Potjer, S. (2018), 'Citizen-generated open data: an explorative analysis of 25 cases', *Government Information Quarterly*, Vol. 35, No 4, Elsevier, pp. 613–621. <https://doi.org/10.1016/j.giq.2018.10.004>
- Notermans, V. I., Montanari, C., Janssen, A., Hölscher, K., Wittmayer, J. M. and Passani, A. (2022), 'Recommendations to mainstream citizen science in policy', *ACTION Project*, Zenodo. <https://doi.org/10.5281/zenodo.5772236>
- Patel, R., Peppin, A., Pavel, V., Brennan, J., Parker, I. and Safak, C. (2021), *Participatory data stewardship: a framework for involving people in the use of data*, Ada Lovelace Institute.
- Ponti, M. and Craglia, M. (2020), *Citizen-generated data for public policy*, Joint Research Centre. <https://ec.europa.eu/jrc/communities/en/community/citizensdata/document/citizen-generated-data-public-policy>
- Shirk, J. L., Ballard, H. L., Wilderman, C. C., Phillips, T., Wiggins, A., Jordan, R., McCallie, E., Minarchek, M., Lewenstein, B. V., Krasny, M. E. and Bonney, R. (2012), 'Public participation in scientific research: a framework for deliberate design', *Ecology and Society*, Vol. 17, No 2, The Resilience Alliance. <https://doi.org/10.5751/ES-04705-170229>
- Sieber, R. E. and Johnson, P. A. (2015), 'Civic open data at a crossroads: dominant models and current challenges', *Government Information Quarterly*, Vol. 32, No 3, Elsevier, pp. 308–315. <https://doi.org/10.1016/j.giq.2015.05.003>
- Suman, A. B., Schade, S. and Abe, Y. (2020), 'Exploring legitimization strategies for contested uses of citizen-generated data for policy', *Journal of Human Rights and the Environment*, Vol. 11, No 3, Edward Elgar Publishing, pp. 74–102. <https://www.elgaronline.com/view/journals/jhre/11-3/jhre.2020.03.04.xml>
- Watson, D. and Floridi, L. (2018), 'Crowdsourced science: sociotechnical epistemology in the e-research paradigm', *Synthese*, Vol. 195, No 2, Springer Nature, pp. 741–764. <https://doi.org/10.1007/s11229-016-1238-2>
- Wiggins, A. and Crowston, K. (2011), 'From conservation to crowdsourcing: a typology of citizen science', *2011 44th Hawaii International Conference on System Sciences*, Institute of Electrical and Electronics Engineers, pp. 1–10. <https://doi.org/10.1109/HICSS.2011.207>
- Wilkinson, M. et al. (2016), 'The FAIR guiding principles for scientific data management and stewardship', *Scientific Data*, Vol. 3, Springer Nature. <https://doi.org/10.1038/sdata.2016.18>

A. Reviewed open data portals

For each of the selected data portals, we describe the initial list of keywords used to identify CGD datasets, which sometimes has produced no results and have required a more tedious manual exploration. We then list the final list of datasets that have been identified and annotated according to our proposed set of dimensions.

A.1. National: Czechia (data.gov.cz)

Used keywords: data 'generovaná občany' (no results), 'občanská věda' (no results), 'účast občanů', 'občané'.

Further processing: A further manual processing needed to be done, reviewing approximately 30 datasets.

Resulting list of datasets:

1. Cizinci podle státního občanství, věku a pohlaví – rok 2004
2. Cizinci podle státního občanství, věku a pohlaví – rok 2005
3. Cizinci podle státního občanství, věku a pohlaví – rok 2006
4. Cizinci podle státního občanství, věku a pohlaví – rok 2007
5. Cizinci podle státního občanství, věku a pohlaví – rok 2008
6. Cizinci podle státního občanství, věku a pohlaví – rok 2009
7. Cizinci podle státního občanství, věku a pohlaví – rok 2010
8. Cizinci podle státního občanství, věku a pohlaví – rok 2011
9. Cizinci podle státního občanství, věku a pohlaví – rok 2012
10. Cizinci podle státního občanství, věku a pohlaví – rok 2013
11. Cizinci podle státního občanství, věku a pohlaví – rok 2014
12. Cizinci podle státního občanství, věku a pohlaví – rok 2015
13. Cizinci podle státního občanství, věku a pohlaví – rok 2016
14. Cizinci podle státního občanství, věku a pohlaví – rok 2017
15. Cizinci podle státního občanství, věku a pohlaví – rok 2018
16. Cizinci podle státního občanství, věku a pohlaví – rok 2019
17. Cizinci podle státního občanství, věku a pohlaví – rok 2020
18. Čestní občané
19. Doprava
20. Pocitová mapa 2016
21. Pocitová mapa horka 2017
22. Životní prostředí
23. Žádosti o vydání občanského průkazu bez strojově čitelných údajů
24. Žádosti o vydání elektronického občanského průkazu se strojově čitelnými údaji a s kontaktním elektronickým čipem
25. Žádosti o pronájem bytů HMP

26. Seznam čestných občanů MČ Praha 11
27. Seznam čestných občanů
28. Dopravní průzkumy pražské MHD
29. Konjunkturální průzkumy
30. Průzkum maloobchodní sítě města Brna

A.2. National: France (data.gouv.fr)

Used keywords: 'Données générées par les citoyens', 'science citoyenne', 'participation citoyenne', 'citoyens'.

Further processing: A further manual processing needed to be done, reviewing approximately 100 datasets.

Resulting list of datasets:

1. Nombre de personnes rickrollées sur data.gouv.fr
2. Contributions à la consultation 'Plan d'action gouvernement ouvert 2017- 2019: idéation'
3. Contributions anonymisées – Appel à idées sur la mise à disposition de données publiques (ouvertes ou en accès restreint)
4. Données essentielles de la commande publique transmises via le PES Marché
5. GeoZones
6. PIAF – Le dataset francophone de Questions-Réponses
7. Programme Entrepreneurs d'Intérêt Général
8. OSM and its reutilisations
9. Grand Débat Participation Citoyenne MEL Contributions libres
10. Grand Débat – Réponses aux questions du Président – territoire MEL
11. Avignon – Budget Participatif 2017
12. Avignon – Budget Participatif 2018
13. Avignon – Budget Participatif 2017
14. Avignon – Végétalisation Participative
15. Projets laureats du budget participatif – Orléans Métropole
16. Conseils participatifs – Mulhouse

A.3. National: Poland

Used keywords: Not applicable.

Further processing: None

Resulting list of datasets: None.

A.4. National: Spain (datos.gob.es)

Used keywords: 'datos generados por ciudadanos', 'ciencia ciudadana', 'participación ciudadana', 'ciudadanos'.

Further processing: A further manual processing needed to be done, reviewing more than a thousand datasets.

Resulting datasets: More than a thousand datasets that satisfy the keywords listed above have been found. This represents 20 % of the total dataset of the portal (57 000 total datasets). Therefore, it will not list any datasets.

A.5. National: United Kingdom (data.gov.uk)

Used keywords: 'citizen-generated data', 'citizen science', 'citizen participation', 'citizens'.

Further processing: A further manual processing needed to be done, reviewing approximately 10 datasets.

Resulting list of datasets:

1. Trafford Council – Allotments
2. Minecraft Cambridge
3. 1992 – 2012 Centre for Environment, Fisheries and Aquaculture Science (Cefas) Seawater temperature records for the UK Shelf – 10 – Citizen Science Diver Recorded Temperatures
4. 2011 Marine Biological Association of the UK (MBA) Batten Bay Corella eumyota survey
5. 2016 Marine Biological Association of the UK (MBA) Batten Bay Corella eumyota survey
6. Woodland sites
7. Records of leaf damage caused by and parasitism of *Cameraria ohridella* in Britain in 2010 collected with a citizen science approach, plus validation of the data
8. 2015 – 2017 Community Seagrass Initiative (CSI) South Coast Cornwall, Devon and Dorset Seagrass Transect Survey

A.6. Regional: Catalonia (governobert.gencat.cat)

Used keywords: 'dades generades per ciutadans', 'ciència ciutadana', 'participació ciutadana', 'ciutadans'.

Further processing: A further manual processing needed to be done, reviewing approximately 400 datasets.

Resulting list of datasets:

1. Estadístiques dels 25 webs més visitats de la Generalitat de Catalunya

2. Avaluació dels tallers de debat dels processos participatius de la Generalitat de Catalunya 2008-2019
3. Registre d'Estudis d'Opinió del Centre d'Estudis d'Opinió
4. Suggestiments, agraïments i queixes rebuts a la Direcció General de la Policia en relació amb l'activitat de la Policia de la Generalitat – Mossos d'Esquadra
5. Dades de trucades operatives gestionades pel CAT112
6. Dades d'incidents operatius gestionats pel CAT112

A.7. Regional: Trent (dati.trentino.it)

Used keywords: 'dati generati dai cittadini', 'scienza dei cittadini', 'partecipazione dei cittadini', 'cittadini'.

Further processing: A further manual processing needed to be done, reviewing more than a thousand datasets.

Resulting list of datasets:

1. Patti di collaborazione (Open data)
2. Serie storica andamento (per argomento) delle comunicazioni inoltrate all'URP (open data)

Given the huge number of matches found, listing all datasets would be equivalent to listing approximately 75 % of the entire portal.

A.8. Local: City of Barcelona (opendata-ajuntament.barcelona.cat)

Used keywords: 'datos generados por ciudadanos' (no results), 'ciencia ciudadana' (no results), 'participación ciudadana' (no results), 'citizen science' (no results), 'citizen-generated data', 'encuesta'.

Further processing: A further manual processing needed to be done with the goal of reviewing the 10 datasets obtained during the keyword search.

Resulting list of datasets:

1. Encuesta de Servicios Municipales de la ciudad de Barcelona – Evolución
2. Encuesta sobre la brecha digital en la ciudad de Barcelona
3. Encuesta de Bienestar Subjetivo de la Infancia en la ciudad de Barcelona
4. Incidencias, quejas, sugerencias, consultas y agradecimientos gestionados por el Ayuntamiento de Barcelona

A.9. Local: City of Berlin (daten.belin.de)

Used keywords: 'von Bürgern erzeugte Daten' (no results), 'Bürgerwissenschaft' (no results), 'Bürgerbeteiligung', 'Bürger'.

Further processing: A further manual processing needed to be done, reviewing approximately 15 datasets.

Resulting list of datasets:

1. Gieß den Kiez – Nutzungsdaten
2. Bevölkerungszahlen Berlin

A.10. Local: City of Dublin (data.smartdublin.ie)

Used keywords: 'citizen-generated data', 'citizen science' (no results), 'citizen participation', 'citizens'.

Further processing: A further manual processing needed to be done, reviewing approximately 25 datasets.

Resulting list of datasets:

1. FixYourStreet.ie reports
2. Customer service requests log DCC
3. Public Realm Street Engagement DCC
4. Freedom of Information Requests DCC
5. Local Elections 2014 Results
6. Webcasting Viewers

A.11. Local: City of Helsinki (hri.fi)

Used keywords: 'kansalaisten tuottama tieto', 'kansalaistiede' (no results), 'kansalaisten osallistuminen' (no results), 'kansalaiset'.

Further processing: A further manual processing needed to be done, reviewing approximately 130 datasets.

Resulting list of datasets:

1. HELSINGIN KAUPUNGIN YMPÄRISTÖRAPORTTI
2. HELSINGIN SEUDUN LIIKENTEEN (HSL) REITTIOPAS API
3. VANTAAN KAUPUNGIN OSALLISUUSKYSELYN VASTAUKSET
4. ULKOMAAN KANSALAISET HELSINGISSÄ KANSALAIUUDEEN SEKÄ SYN- TYMÄ ALUEEN JA SUKUPUOLEN MUKAAN 2017 ALKAEN

5. JÄTTEIDEN LAJITTELUAKTIIVISUUS PÄÄKAUPUNKISEUDULLA
6. HELSINGIN PYÖRÄILYBAROMETRIKYSÉLYN VASTAUKSET
7. HELSINGIN SEUDUN LIIKENTEEN (HSL) ASIAKASTYYTYVÄISYYS- TUTKIMUS
8. VANTAAN ASUKASBAROMETRIN VASTAUKSET
9. HELSINGIN KAUPUNGIN YRITYSHAASTATTELUTUTKIMUS
10. HELSINGIN JA VANTAAN YMPÄRISTÖASENNEKYSÉLYN VASTAUKSET
11. HELSINGIN TUULIVOIMAKYSÉLY 2015
12. HELSINKI 2050 -KYSÉLYN VASTAUKSET
13. KIVA KESKUSTA KÄVELIJÖILLE -KYSÉLYN VASTAUKSET
14. VANTAAN KAUPUNGIN OSALLISUUSKYSÉLYN VASTAUKSET
15. VANTAAN TULEVAISUUSKYSÉLY 2017
16. HELSINGIN VAPAA-AIKAKYSÉLYN 2018 VASTAUKSET
17. HELSINGIN KANSALLINEN KAUPUNKIPUISTO -KYSÉLYN VASTAUKSET
18. HELSINGIN LIIKENNETURVALLISUUSKYSÉLY ASUKKAILLE
19. AARTEENI HELSINGISSÄ-KYSÉLYN VASTAUKSET
20. HELSINGIN KAUPUNGIN VIRASTOJEN JA LIIKELAITOSTEN PUHE-LINNUMEROT
21. HELSINGIN JOHTAMISEN UUDISTAMINEN -VERKKOKYSÉLYN VASTAUKSET
22. MATKUSTAJA-ALUSTEN AIKATAULU-, LIIKENNÖINTI-, JA SIJAINTI- TIETOJA HELSINGIN JA TALLINNAN VÄLILLÄ
23. VANTAAN KAUPUNGIN PALVELUJEN SAAVUTETTAVUUSKYSÉLY 2018
24. HELSINGIN PYÖRÄVÄYLIEN ONGELMAKOHDEKYSÉLYN VASTAUKSET
25. VANTAAN PYÖRÄLIIKENTEEN KEHITTÄMISKYSÉLY
26. HELSINGIN KAUPUNGIN YMPÄRISTÖRAPORTTI

A.12. Local: City of Lyon (data.grandlyon.com)

Used keywords: 'Données générées par les citoyens', 'science citoyenne', 'participation citoyenne', 'citoyens'.

Further processing: A further manual processing needed to be done, reviewing approximately 72 datasets.

Resulting list of datasets:

1. Annuaire pour les professionnels de solutions de réduction et gestion des déchets de la Métropole de Lyon
2. Annuaire pour les consommateurs de solutions de réduction des déchets et d'économie circulaire de la Métropole de Lyon
3. Vie citoyenne de la commune de Lyon
4. Aires d'accueil des gens du voyage à la Métropole de Lyon
5. Points d'intérêt touristiques de la Métropole de Lyon

6. Consommation annuelle électricité/gaz par iris et par code NAF dans la Métropole de Lyon

A.13. Local: City of London (data.london.gov.uk)

Used keywords ⁽⁸⁾: 'citizen-generated data', 'citizen science' (no results), 'citizen participation', 'citizens'.

Further processing: A further manual processing needed to be done, reviewing approximately 10 datasets.

Resulting list of datasets:

1. European quality of life
2. Community strength indicators

A.14. Local: City of Madrid (datos.madrid.es)

Used keywords: 'datos generados por ciudadanos' (no results), 'ciencia ciudadana' (no results), 'participación ciudadana', 'ciudadanos'.

Further processing: A further manual processing needed to be done, reviewing approximately 500 datasets.

Resulting list of datasets:

1. Participación Ciudadana. Debates y Propuestas
2. Participación Ciudadana. Presupuestos participativos
3. Participación Ciudadana. Consultas Ciudadanas
4. Consultas urbanísticas comunes y especiales
5. Encuesta de calidad de vida y satisfacción con los servicios públicos de la ciudad de Madrid (serie)
6. Encuesta de satisfacción de usuarios de la Oficina de Atención al Ciudadano de Madrid Salud (Serie)
7. Sondeo de la valoración y concienciación ciudadana sobre limpieza urbana en la ciudad de Madrid (serie)
8. Sondeo de opinión ciudadana sobre la percepción de seguridad en el distrito de Villaverde
9. Sondeo sobre la Consulta Ciudadana organizada por el Ayuntamiento de Madriden febrero de 2017
10. Cita Previa (Línea Madrid)
11. Atención personalizada en Línea Madrid (canal presencial, telefónico y telemático @lineamadrid)

⁽⁸⁾ The portal does not provide any tool or map, nor does it provide a query engine to search datasets. So it is one of the worst portals to find concrete datasets.

12. Avisos ciudadanos
13. Avisos y reclamaciones recepcionados en las secciones de bicicleta pública (BiciMAD), señalización y SER
14. BiciMAD. Datos de los itinerarios de las bicicletas eléctricas

A.15. Local: City of Milan (dati.comune.milano.it)

Used keywords: 'dati generati dai cittadini', 'scienza dei cittadini', 'partecipazione dei cittadini', 'cittadini'.

Further processing: A further manual processing needed to be done, reviewing approximately 35 datasets.

Resulting list of datasets:

1. Case dell'acqua nel Comune di Milano
2. Fontanelle nel Comune di Milano
3. OpenWifiMilano (multiple)
4. Bilancio Partecipativo 2017/2018
5. Referendum 2020 (multiple)
6. Elezioni (multiple)
7. Customer satisfaction (multiple)
8. Consumi delle famiglie: spesa media mensile per categoria di di spesa e classedi reddito del nucleo familiare

A.16. Local: City of Zaragoza (zaragoza.es/sede/portal/datos-abiertos)

Used keywords: 'datos generados por ciudadanos' (no results), 'ciencia ciudadana' (no results), 'participación ciudadana', 'ciudadanos'.

Further processing: A further manual processing needed to be done, reviewing approximately 160 datasets.

Resulting list of datasets:

1. Estaciones y uso de Bizi Zaragoza
2. Incidencias en la Vía publica
3. Anuncios para Jovenes del CIPAJ
4. Zaragoza sin barreras
5. Agenda de Zaragoza
6. Estado del Trafico
7. Servicio de quejas y sugerencias

8. Mapas colaborativos
9. Encuesta de satisfacción de los servicios municipales
10. Datos de uso de la tarjeta Interbus
11. Indicadores de tráfico y movilidad
12. Encuesta Condiciones de vida, percepción y valoración emocional de la ciudadanía durante el confinamiento