

data.
europa.
eu
academy 

Citizen-generated data on data.europa.eu

21-04-2022



data.europa.eu The official portal
for European data



Agenda for the webinar

- Introduction to citizen-generated data (CGD)
- The research method followed for this report
- Results
 - Dimensions to categorise citizen-generated data
 - Analysis of the field study
- Conclusions and recommendations
- Next steps, including a survey to identify candidates for interviews and workshops
- Feedback and closing

1. Introduction

Citizen-Generated Data:

What is it? Why is it important in the European context?

Where do we start from?

- **Citizens generate large amounts of data** as part of their daily activities, consciously or unconsciously
 - Often managed by private companies to improve their services (e.g., business ratings in Google Places and FourSquare)
 - Or made available as open data on dedicated sites (e.g., FixMyStreet, Wikidata, OpenStreetMap).
 - Or coming from specific Citizen Science projects
 - Or generated as part of their interactions with public administrations (claims and suggestions, public participation, transport card usage)
- **These types of data sources are not widely available in data.europa.eu (nor in other open data portals) by design**, as these portals usually limit themselves to government-generated data.
- **Citizen-sourced data would complement existing official datasets** (e.g. OpenStreetMap may have more up-to-date, detailed information about cities than city authorities) while generating more awareness about data.europa.eu among the public.

Goal of this series of reports

Our main Research Questions

- Which opportunities and challenges are associated with the inclusion of citizen-generated data in data.europa.eu?
- Would traditional data providers be willing to incorporate such data sources, or allow derivations of their own data sources with citizen-generated data?
- How could we promote citizens to become active data providers for data.europa.eu? (if it makes sense)

Question time (2 minutes)...

**What is CGD?
Is it available in European Open Data
Portals?**

Feel free to use the chat to provide your initial informal view on this

Some (partial) definitions found in the literature

- **Sieber & Johnson, 2015:** *important form of open data for public governance, as it facilitates collaborative action by various stakeholders to tackle public concerns.*
- **Ponti & Craglia, 2020:** *data that **people or their organisations produce** to directly monitor, demand or drive change on issues that affect them.*
- **DataShift, 2015:** *is **actively given by citizens**, providing direct representations of their perspectives and an alternative to datasets collected by governments or international institutions.*
- **Suman et al., 2020:** *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**.*

So..., What is Citizen-Generated data?

- **Multiple definitions:** most of them differ on the required level of consciousness in the data production process to be considered CGD.
- **Multiple Applications:** CGD is a massive source of value that may be exploited by public and, mainly, private companies.
- **Social Implications:** public participation promotes democratic health and enhances social involvement in public projects.

*“The data that individuals consciously **and unconsciously** generate **or curate** and that are openly available for use in the public domain **(via an open data portal)**”*

2. Research methodology

How can we address these questions?

Our research methodology

1. **Desk research (based on a Systematic Literature Review protocol):** Identify CGD definitions, discover how CGD has been used by public agencies (e.g., improve data quality, enhance citizen participation), and identify common dimensions to categorise CGD (in an open data context).
 - **Sources used:** Web of Science and Joint Research Center (JRC) Publications Repository
 - **Topics searched:** *citizen-generated data, citizen science, open science, open data*
 - **Papers found:** 146 Web of Science papers and 19 JRC Science Hub
 - **Inclusion/Exclusion criteria:** (1) It contains clear references to the publication of CGD as open data; and (2) it discusses potential categories or dimensions to classify or characterise CGD. → 7 resulting papers (+1 using snowballing via their references)

Our research methodology

- 2. Definition of CGD dimensions:** to deal with variability in definitions and systematically classify CGD in European Open Data Portals.
- 3. Analysis of a sample of European Open Data Portals (national, regional, local):** to find which types of CGD are offered there, and which tools are integrated to ease citizen participation in data provisioning and curation.
- 4. Today 's Webinar:** knowledge sharing, initial feedback and new ideas, and initial steps for stakeholder identification for the next phases.

3. Results (I): dimensions to categorise CGD

How can we classify Citizen-Generated Data in a meaningful way?

Proposed dimensions to classify CGD

Data Portal (global) Dimensions

- Percentage of CGD datasets in the portal
 - Percentage (%)
- Existence of general guidelines for publishing CGD
 - (y)es / (n)o
- Availability of methods and tools for CGD management and use
 - (y)es / (n)o
- Proposal of quality assurance mechanisms for CGD
 - (y)es / (n)o

Proposed dimensions to classify CGD

Dataset-specific Dimensions

- Area
 - Geolocation (gl), Environmental (e), Citizen Science (cs), Votes (v), Surveys (su), Questions and answers (qa), statistics (st).
- Actor roles
 - Citizens (c), Public administration (pa), Others (o)
- **Primary / Secondary CGD**
- Data **by** or **about** citizens
- Specific CGD Dataset Guidelines
 - (y)es / (n)o
- Expected Policy/Operational Impact
 - (y)es / (n)o
- Data Format
 - xml, json, csv, xlsx, zip or others
- Licensing
 - Open/Closed

Some illustrative examples

- Estaciones y uso de Bizi (Zaragoza) → Public bicycles
 - Geolocalisation (gl) / Funded and Initiated by Public administration (pa) / secondary
 - <https://www.zaragoza.es/sede/servicio/catalogo/70>
- Incidencias, quejas, sugerencias, consultas y agradecimientos (Barcelona) → Questions/Answers/Suggestions
 - Survey (su) / Funded and Initiated by Public administration (pa) / primary
 - <https://opendata-ajuntament.barcelona.cat/data/es/dataset/iris>
- Avignon - Budget Participatif 2017 (France) → Participatory budgets
 - Questions and Answers (qa) / Funded and Initiated by Public administration (pa) / primary
 - <https://www.data.gouv.fr/fr/datasets/avignon-budget-participatif-2017-projets-soumis-au-vote-et-retenus/>
- KIVA KESKUSTA KÄVELIJÖILLE -KYSELYN VASTAUKSET (Helsinki) → Survey about visitor center
 - Survey; Question and Answers (qa) / Funded and Initiated by Public administration (pa) / primary
 - <https://hri.fi/data/fi/dataset/kiva-keskusta-kavelijoille-kyselyn-vastaukset>

4. Results (II): field analysis of open data portals

What have we found in the European context?

Analysed portals

National

1. Czechia
2. France
3. Poland
4. Spain
5. United Kingdom

Regional

1. Catalonia, Spain
2. Province of Trento, Italy

Local

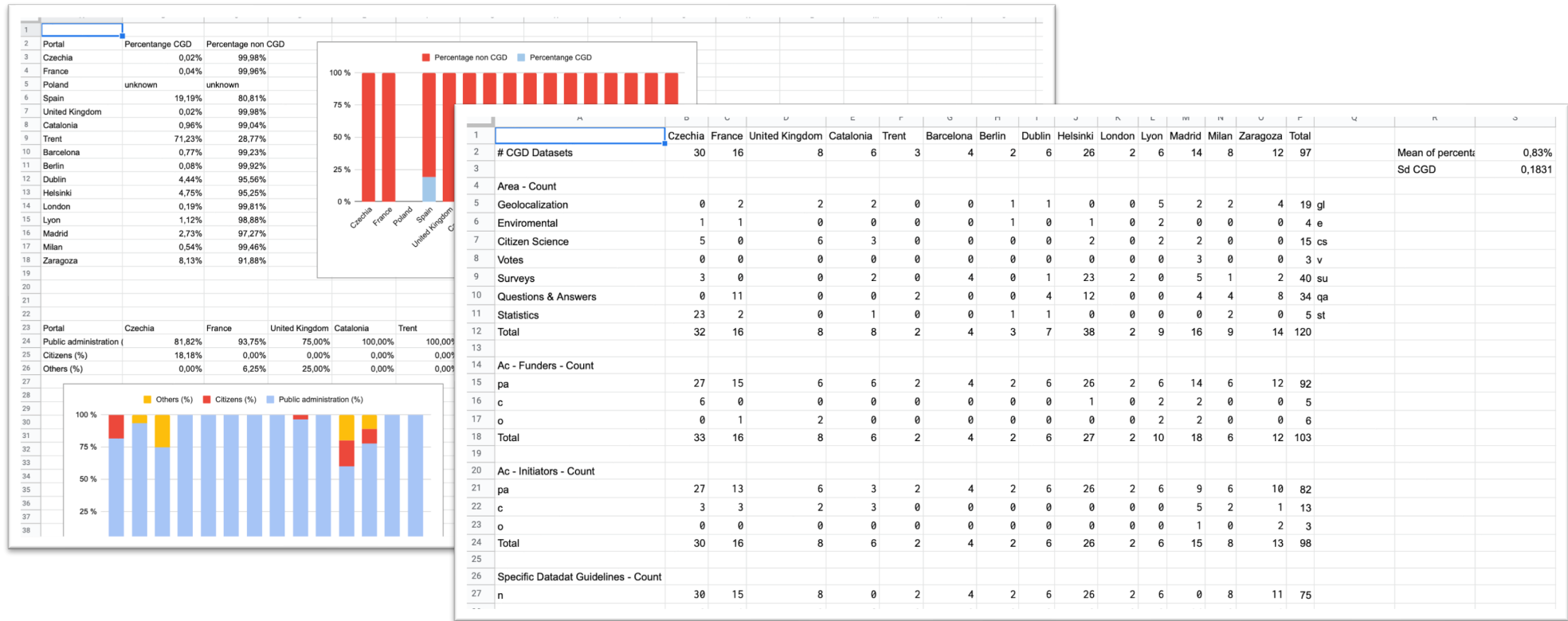
1. Barcelona, Spain
2. Berlin, Germany
3. Dublin, Ireland
4. Helsinki, Finland
5. London, UK
6. Lyon, France
7. Madrid, Spain
8. Milan, Italy
9. Zaragoza, Spain

We will be surely missing relevant cases (e.g., the city of Valencia, in Spain, released a strategy for incorporating CGD in March 2022)

Use the survey later to give us some more examples, or contact us by e-mail

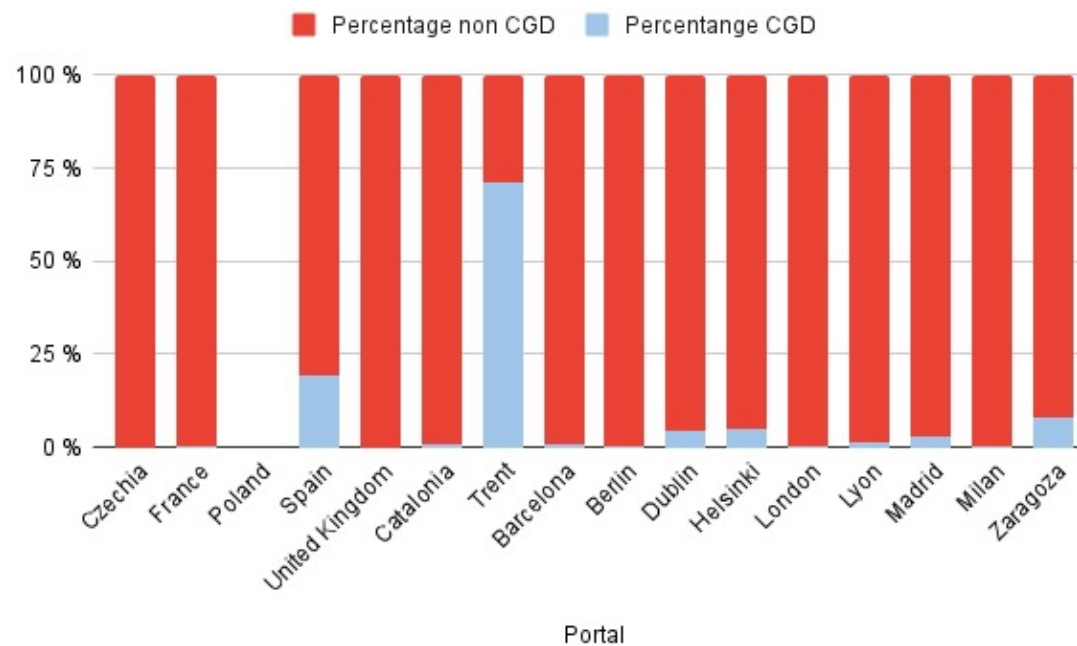
Results

Spreadsheets for the annotation process



Results

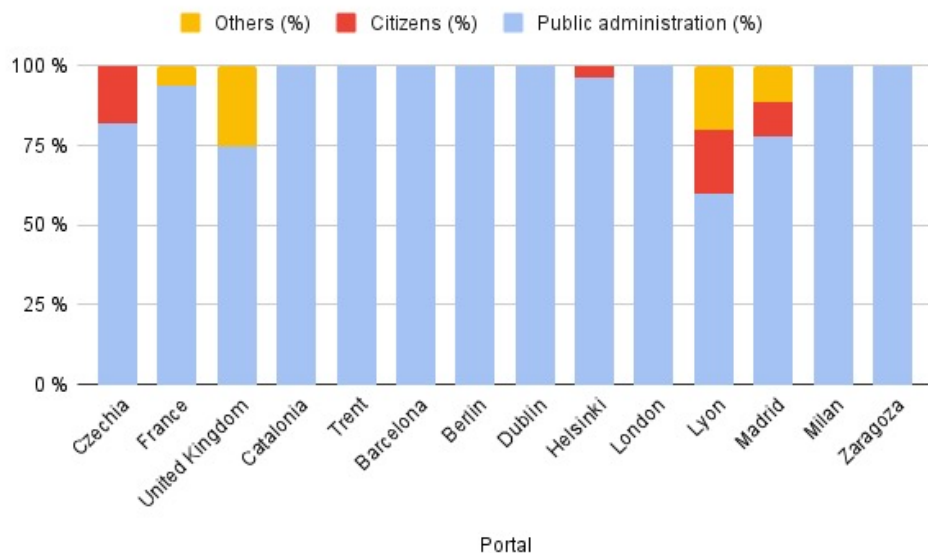
Percentage of CGD in each of the studied portals in relation to all datasets that are published



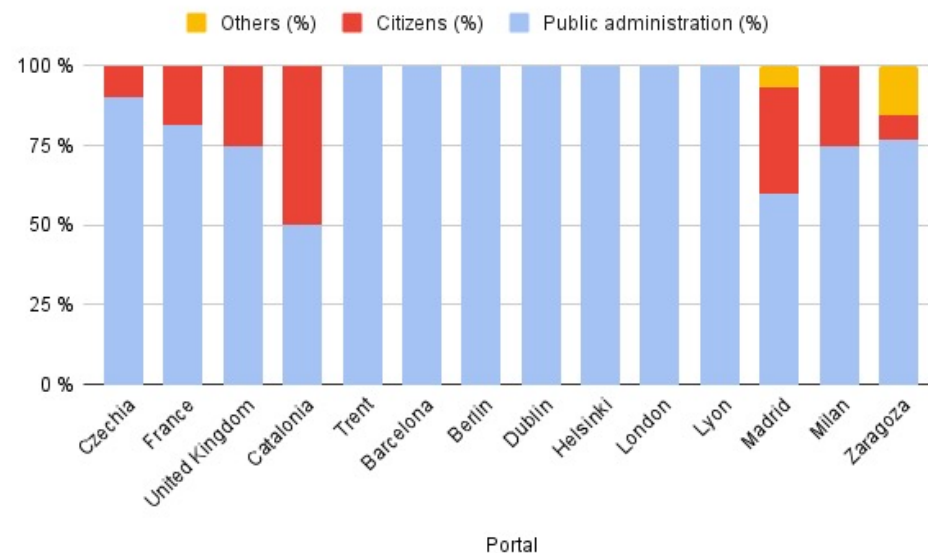
Results

Proportion of actors involved in CGD published in open data portals.

As project funders

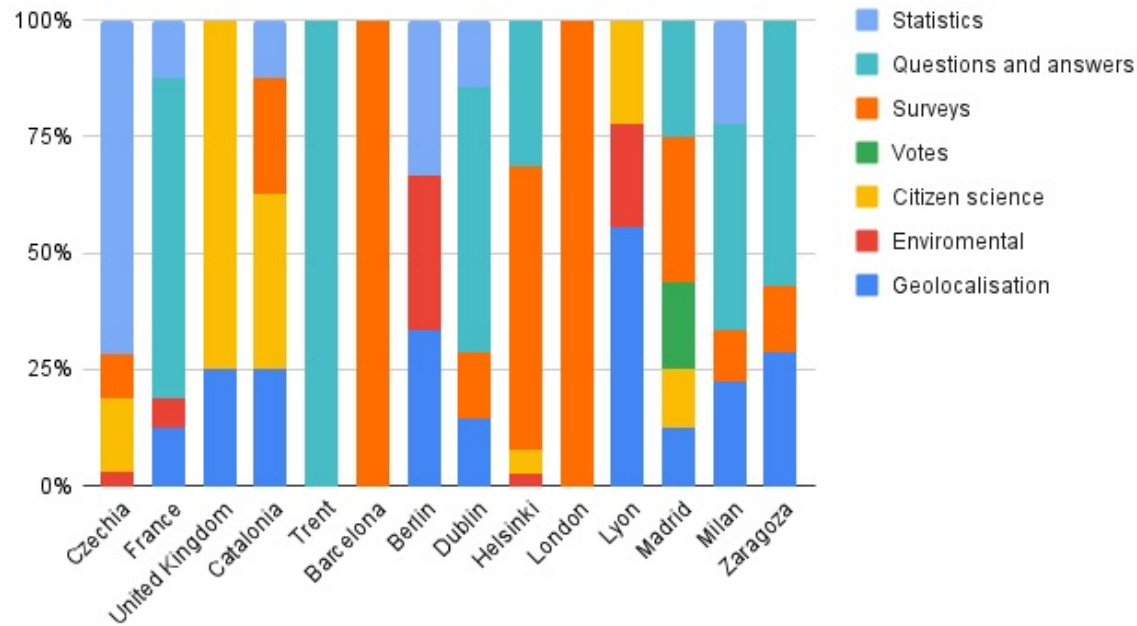


As project initiators



Results

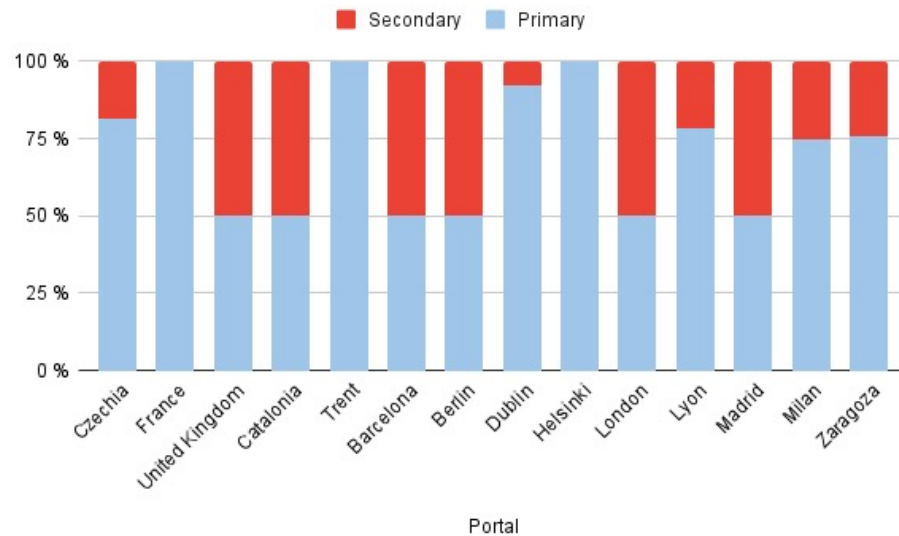
CGD datasets per Portal and Area



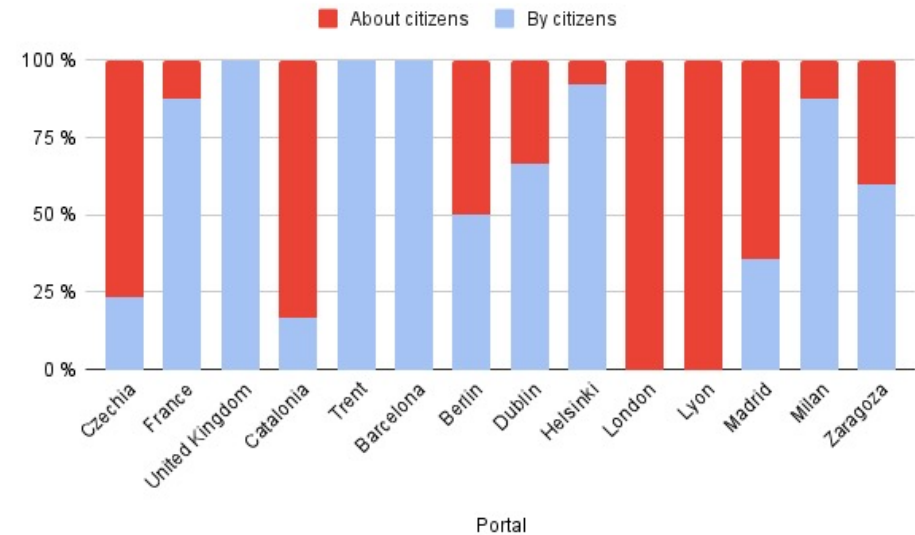
Results

Citizen's role in CGD published in open data portals.

Primary/Secondary CGD

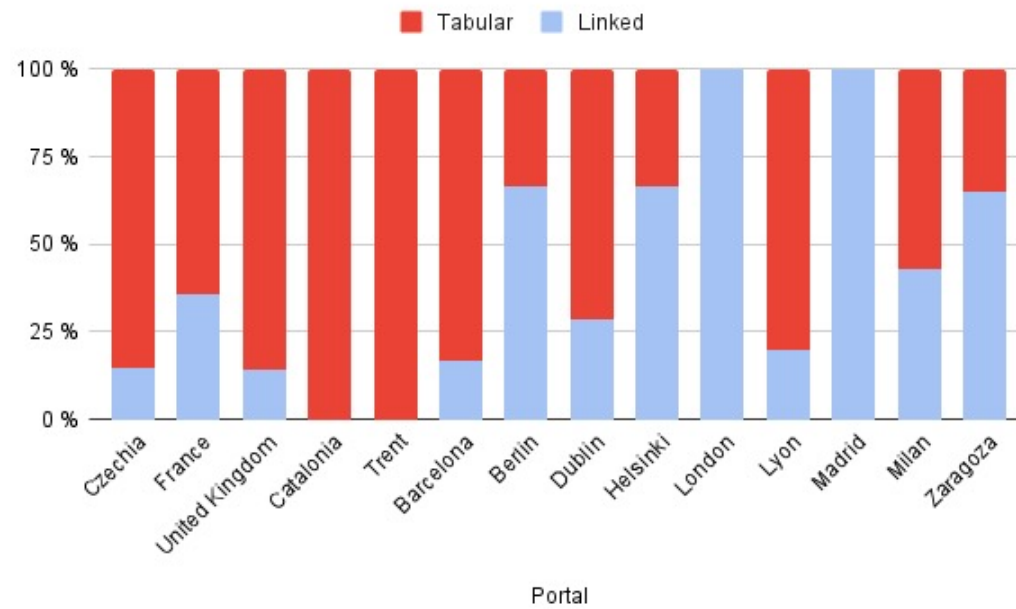


By/About CGD



Results

Formats of CGD datasets



5. Conclusions and Recommendations

Some final ideas.

Conclusions

1. In general, the presence of CGD in open data portals is very low
2. For existing CGD published as open data, citizens are mostly involved in generating or curating data. The remaining work required to publish the data is driven by public administrations.
3. “Questions and Answers”, “Surveys” and “Statistics” are the most common areas of CGD
4. Primary CGD is more common than secondary CGD
5. Most of CGD is shared with open licenses.

Conclusions

6. Most of CGD datasets are available in open formats
7. None of the studied portals included documentation about how to contribute and use CGD
8. In general terms, portals do not offer tools to facilitate citizen contribution
9. We could find no evidence of participatory approaches to design data pipelines or collect and implement feedback from citizens on a broader data strategy
10. No general guidelines on how to govern CGD

Initial recommendations

1. Actively seek valuable CGD assets through open calls and partnerships with key citizen-science players (e.g., ECSA, national and regional Citizen Science offices, Citizen Science projects)
2. 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"
3. Include keywords/tags in EU official languages to facilitate comparative studies using multiple datasets
4. Establish procedures to capture CGD processes and data validation methods, so as to increase trust of third-party data users
5. Extend data and metadata quality metrics with those specific to CGD

Initial recommendations

6. Include CGD aspects in upcoming open data reports, such as the European Analysis of Open Data Portals
7. Collect new and tag existing use cases from data.europa.eu to showcase the value of CGD datasets
8. 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
9. Create tools and applications that consume this type of data and allow citizens to contribute to the original data sources
10. 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

6. Next steps

Next steps

Two more related reports expected in the coming months:

- Identification of other types of citizen-generated data sources (2022 Q3)
 - Based on a workshop with a selected set of participants
 - Fill in the following survey if you want to participate (or contact us directly)
 - <https://ec.europa.eu/eusurvey/runner/CitizenGeneratedDataSurvey>
 - Main objective: which types of citizen-generated data sources may be published as open data by data providers, including data.europa.eu?
- A model for the governance of citizen-generated data, with three cases (2023 Q1)
 - Based on interviews with stakeholders from at least three case studies

7. Feedback and Closing

Please
provide us
your
feedback!





data.
europa.
eu
academy 

Thank you very much!

info@data.europa.eu



 The official portal
for European data

