UNIVERSITY OF TARTU
Faculty of Science and Technology
Institute of Computer Science
Software Engineering Curriculum

Fillip Molodtsov

An Integrated Usability Framework for Evaluating Open Government Data Portals: Comparative Analysis of EU and GCC Countries

Master's Thesis (30 ECTS)

Supervisor(s): Anastasija Nikiforova, PhD

Tartu 2024

An Integrated Usability Framework for Evaluating Open Government Data Portals: Comparative Analysis of EU and GCC Countries

Abstract: This study explores the critical role of open government data (OGD) portals in fostering transparency and collaboration between diverse stakeholders. Recognizing the challenges of usability, communication with diverse populations, and strategic value creation, this paper develops an integrated framework for evaluating OGD portal effectiveness that accommodates user diversity (regardless of their data literacy and language), evaluates collaboration and participation, and the ability of users to explore and understand the data provided through them. The framework is validated by applying it to 33 national portals across European Union (EU) and Gulf Cooperation Council (GCC) countries, as a result of which each portal was assessed, statistics about the portal performances were gathered, portals were ranked, best practices and pain points for the portals were derived, trends in portal design and collaborative initiatives between portals were identified, cluster analyses based on the score matrix were conducted. Cluster analyses (K-Means, hierarchical) determine which clusters to consult depending on the dimensions' strengths. Nineteen recommendations have been made based on all the above-mentioned analyses. Notably, the study unveils the competitive and innovative nature of GCC OGD portals, pinpointing specific areas such as multilingual support and data understandability. The findings underscore the growing trend of exposing data quality metrics and advocate for enhanced communication channels between users and portal representatives. Overall, the study contributes to accelerating the development of user-friendly, collaborative, and sustainable OGD portals while addressing gaps identified in previous research.

Keywords:
Open data portal, Open data, Open government data portal, Open government data, OGD portal, Framework, Usability, Sustainability, Open data ecosystem, European Union, GCC, Gulf Cooperation Council

CERCS: T120 - Systems engineering, computer technology P170 - Computer science, numerical analysis, systems, control P175 - Informatics, systems theory

Integreeritud kasutatavuse raamistik avatud valitsuse andmeportaalide hindamiseks: ELi ja Pärsia lahe koostöönõukogu riikide võrdlev analüüs

Lühikokkuvõte: See uuring uurib avatud valitsuse andmete (OGD) portaalide kriitilist rolli läbipaistvuse ja koostöö edendamisel erinevate sidusrühmade vahel. Tunnustades kasutatavuse, mitmekesise elanikkonnaga suhtlemise ja strateegilise väärtuse loomise väljakutseid, töötab käesolev dokument välja integreeritud raamistiku OGD portaali tõhususe hindamiseks, mis mahutab kasutajate mitmekesisust (sõltumata nende andmekirjaoskusest ja keelest), hindab koostööd ja osalemist ning kasutajate võimet uurida

ja mõista nende kaudu esitatud andmeid. Raamistiku valideerimiseks kohaldatakse seda 33 Euroopa Liidu (EL) ja Pärsia lahe koostöönõukogu (GCC) riigi portaali suhtes, mille tulemusena hinnati iga portaali, koguti statistikat portaalide toimivuse kohta, reastati portaalid, tuletati portaalide parimad tavad ja valupunktid, selgitati välja portaalide kujundamise suundumused ja koostööalgatused portaalide vahel, tehti punktimaatriksil põhinev klasteranalüüs. Klastrite analüüs (K-Means, hierarhiline) määrab, milliste klastritega konsulteerida, sõltuvalt dimensioonide tugevustest. Kõigi eespool nimetatud analüüside põhjal on esitatud üheksateist soovitust. Eelkõige paljastab uuring GCC OGD portaalide konkurentsivõime ja uuenduslikkuse, tuues välja konkreetsed valdkonnad, nagu mitmekeelne tugi ja andmete arusaadavus. Tulemused rõhutavad andmete kvaliteedi mõõdikute paljastamise kasvavat suundumust ning toetavad kasutajate ja portaalide esindajate vaheliste sidekanalite täiustamist. Üldiselt aitab uuring kiirendada kasutajasõbralike, koostööl põhinevate ja jätkusuutlike OGD-portaalide arengut, kõrvaldades samas varasemates teadusuuringutes tuvastatud lüngad.

Võtmesõnad:
Open data portal, Open data, Open government data portal, Open government data, OGD portal, Framework, Kasutatavus, Jätkusuutlikkus, Avatud Data Ecosystem, Euroopa Liit, GCC, Pärsia lahe koostöönõukogu

CERCS: T120 - Süsteemitehnoloogia, arvutitehnoloogia P170 - Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine (automaatjuhtimisteooria) P175 - Informaatika, süsteemiteooria

Contents

1 Introduction

A fundamental principle of open government is transparency of government information and operations, which serves as the basis for its two pillars: *participation* and *collaboration* [Oba09]. Open data and Open Government Data (OGD) are a demonstration of transparency, which can create new opportunities for participation and interaction with government, and also offers new grounds for collaboration between diverse stakeholders [SZJG15].

Open government data is the practice of making government data available to the public for use in response to public demand for access and reuse [ZJD15]. Municipal, state, federal, and national entities are becoming data publishers. Users, including citizens, businesses, and especially Small and Medium-sized Enterprises (SME), media (journalists) participating in the Open Government Data (OGD) movement, create applications, services, maps and other OGD re-use outputs. It also improves data, generates articles and news, turning government data into information and actionable insights of public value, often of social, environmental, and economic [SGJ15, PZJ20c, GPC23]. Furthermore, OGD efforts are propelled by legislative requirements [Mad] and international agreements such as the Open Government Partnership (OGP) [NM21].

Yet the OGD movement has experienced a decline in popularity. As a result, today, at risk of a "data winter" with limited balancing of interests of private entities and broader societal demands for data accessibility, there is an urgent need to create a data ecosystem in which data is not a commodity to be traded, but a resource to empower communities and science and contribute building a more informed and equitable world [Ver24], where the OGD and open data portal can be an asset for this since it becomes the primary gateway for data users to access OGD [JCZ12]. According to research [ZF19], OGD portals have the potential to foster business opportunities (especially in the context of SME) and civic engagement. The goal is to make interacting with the portal as simple and user-friendly as possible so that everyone, regardless of age, gender, digital literacy or data literacy level, or background, can benefit from the data accessed through it. To this end, the portal should be convenient, i.e., user-friendly, and suitable for all potential users regardless of their expertise with OGD [WRC18, NL21, NM21, MAL$^+$21].

Portal usability, communication with diverse populations, and strategic value creation remain the biggest challenges for OGD portals [SZC$^+$15, Car22, Nik20b, MTKK18]. The European Commission's Open Data Maturity Report (ODM report) [Pag23] found that 11 countries (almost half of EU members) scored 90% or above on the portal dimension, with Poland, Estonia, and Ireland receiving the highest scores. However, a high ranking does not imply perfection; it is only relative quality. The ODM reports, however, are based on self-reports provided by representatives of the OGD initiative [Car22], which may raise concerns regarding the credibility of the results and suggests that the user's perspective may be omitted from its assessment as it is not the purpose of this index. Excluding users without domain knowledge from the data ecosystem consideration has been

a common concern in recent years [NTM19]. Problems with OGD initiatives include poor relationships between citizen and government, limited participation, data retrieval issues, poor data quality, lack of standardization, and limited data representation capabilities (raw data, dashboards, statistical reports) [NTM19, Nik20b, MTKK18, SGLPM18].

To accelerate efforts to develop user-friendly, collaborative, robust, and sustainable portals, it is essential to identify the current state of the art and the best practices that those portals should adhere to, taking into account current trends both in the field of open data and more broadly in the fields of software engineering (SE), human–computer interaction (HCI), user experience (UX). This dynamic of continuous development leads to the fact that many developed frameworks become irrelevant or limited, and the evaluation of portals using them and the further development of an agenda to improve their quality prevents the implementation of a sustainable portal that would meet the needs and expectations of users. Many existing indexes and benchmarks have been proposed in the literature to evaluate OGD efforts, including OGD portals [MHL18, LN21, ALM$^+$18, SJ15, ADAD17, MJJ21, ZJD14]. However, [Kao23] that studied many of them, suggests that future research should try to integrate the different frameworks since although the specific target area of OGD benchmark could be performed separately, it would be useful to have an integrated framework. In addition, it is emphasized that benchmarks with a large number of geographic regions are needed. In this study, both gaps are addressed by answering the raised call.

The objective of this study is to develop an integrated framework for evaluating the usability of OGD portals. This framework is based on previous frameworks identified through the Systematic Literature Review (SLR) complemented with criteria defined on the basis of OGD portal weaknesses reported in the literature but not being part of the above frameworks and obtained from portal examinations through desk research. The developed framework is tested on a sample of national portals of the European Union (EU) and Gulf Cooperation Council (GCC) countries. Applying the framework to a sample of 33 countries allows for testing the framework and analyzing portals constituting the sample. The selection of EU national OGD portals as part of the sample reflects the competitiveness and legislative initiatives driving EU portals. In contrast, selecting those from the GCC addresses the gap in understanding the current status of the portals in the region. The framework ranks OGD portals, identifies the most competitive portals, good practices that lower-performing portals can learn from, and common shortcomings. Ultimately, the OGD initiative aims to assist governments in becoming more responsible and transparent in storing, accessing, analyzing, and sharing data. Portals that are more user-friendly and easier to use, potentially increase user engagement. Greater engagement means closer interaction between data sources, producers, and consumers. Tighter and closer interaction and collaboration result in richer portal content, which starts the cycle of open government (data) success, thereby contributing to the maturity, resilience, and sustainability of the open data ecosystem.

The EU and GCC national OGD portals were chosen as the sample for several reasons. EU national portals are among the most competitive in the world due to legislative initiatives and international cooperation [NM21, MAL+21]. In addition, numerous studies have tested their frameworks on EU government portals. Due to the dynamic evolution of OGD portals over the years, ranking results although may be consistent with earlier findings, but with new best practices or recommendations for improving underperforming portals, which corresponds to the recommendations of existing studies to revise the results in the future, taking into account the development of both the portal in question and information and communication technology (ICT) in general. On the other hand, the evaluation of GCC national portals has not been sufficiently studied. Several GCC portals have been covered in previous research, however using the same framework [MHL18], which limits the understanding of their current situation. Due to the rapid development of technology and portals, these analyses and framework can be seen as outdated. Thus, the status of the GCC OGD portals is unknown. According to [Mei19], cooperation between the EU and the GCC countries is crucial to better achieve their political and economic goals. From this point of view, it is important to compare OGD portals from these two regions and determine whether they are comparable and can cooperate ensuring cross-border (and interregional) interoperability and thereby (indirectly) strengthen political and economic ties.

The methodology used in the paper involves the application of Design Science Research (DSR) to develop the integrated usability framework for evaluating OGD portals. The research goes through 5 main stages: information and requirements collection (1), prototyping the framework (2), testing the prototype with portal assessment (3), compiling the framework (4), framework testing, and result analysis (5). As part of the result analysis, cluster analysis, specifically K-Means and hierarchical clustering, is conducted to understand the relationships and recurring patterns among the portals based on their performance metrics.

The paper thus presents an integrated usability framework for evaluating OGD portals. The framework is validated by applying it to 33 national portals across EU and GCC countries, allowing for the ranking of OGD portals and identification of good practices and common shortcomings. The research highlights the competitive and creative character of GCC OGD portals while also indicating areas for development, such as multilingual support and data understandability. Additionally, the need to disclose data quality metrics and advocate for improved communication channels between portal users and representatives is underscored. Based on the analyses of the results, recommendations are formulated.

Theoretically, this study revises and improves criteria and metrics for evaluating the effectiveness and quality of open data portals, incorporates recent literature trends, and sheds light on understudied GCC and EU OGD portals. By evaluating the portals and performing cluster analysis, practical implications are derived in the form of qualitative

and quantitative analyses that portal stakeholders may employ to develop sustainable, collaborative, user-friendly, and robust portals.

The rest of the paper is structured as follows: Section 2 provides the research methodology, Section 3 presents the results of the SLR, Section 4 presents the developed framework, Section 5 presents the results of applying the framework to selected portals along with the cluster analyses based on the score matrix, Section 6 provides a discussion of the results, including a result mapping to prior research, as well as portal recommendations, Section 7 presents the limitations of the study, and Section 8 concludes the paper.

2 Methodology

In order to accomplish the objective of this thesis, which is the creation of an integrated framework for evaluating the usability of an open data portal, the Design Science Research (DSR) methodology [Hev07] has been used. Design Science Research (DSR) is a research paradigm employed in information systems and computer science to address complex problems by designing and evaluating innovative artifacts [Hev07]. Grounded in the principles of rigorous scientific inquiry, DSR integrates problem-solving with the creation of novel solutions, emphasizing the iterative refinement of artifacts through cycles of design, implementation, and evaluation. Central to DSR is the collaboration between researchers and practitioners, ensuring that the developed artifacts contribute to theoretical knowledge and offer practical utility in real-world contexts. In the current case, the collaboration with practitioners is omitted. However, the resulting integrated framework is repeatedly reviewed by 2 experts in the field of open data (OD) portals user experience (UX). A person is considered to be an expert if the person has expertise in computer science and information systems, works with open (government) data and data portals daily, meeting the expert profile according to the derivation of the International Certification of Digital Literacy (ICDL) proposed in [LMV+21]. Figure 1 provides the visual representation of the DSR methodology, where the cycles are illustrated: Relevance, Design, and Rigor. By following this methodology, the integrated framework is developed (during the "Design Cycle") in accordance with the knowledge base (during the "Rigor Cycle") and the environment (during the "Relevance Cycle"), testing the framework against portals.

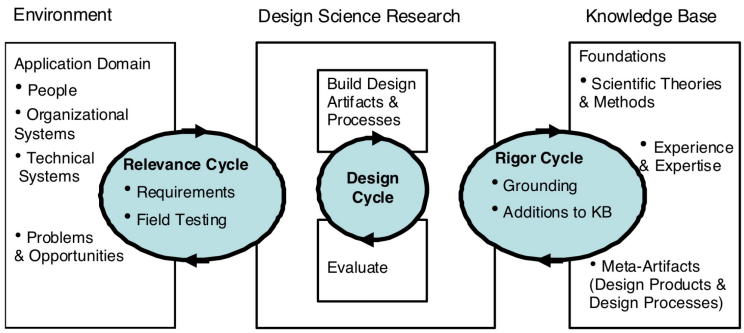


Figure 1. Design Science Research Methodology [Hev07]

The research goes through 5 main stages: information and requirements collection (*Stage 1*, Rigor Cycle), prototyping the framework (*Stage 2*, Design Cycle), conducting portal assessment to test the prototype (*Stage 3*, Relevance Cycle), compiling the framework (*Stage 4*, Design Cycle), framework testing and result analysis (*Stage 5*, Relevance

Cycle). In Stage 1 a systematic literature review (SLR) is carried out. Next, in Stage 2, the draft version of the framework is compiled and preliminary reviewed. In Stage 3, the artifact produced in Stage 2 is tested against a small subset of EU OGD portals to see whether additional changes to the framework should be introduced. In stage 4, suggestions for improvement are considered, the framework is improved, and the final version is reviewed. In Stage 5, the integrated framework is tested on 33 national OGD portals with 27 EU and 6 GCC portals, and the results are analyzed qualitatively and quantitatively, with cluster analysis conducted based on the generated quantitative data. Based on those results and analyses, recommendations are then defined.

2.1 Information and requirements collection

In *Stage1* SLR is carried out, which findings are combined with the comments from leading (highly cited and accepted within the community) experts on the principles of portal design [JCZ12, PZJ20a, PZJ20c, PZJ20b, MJJ21]. For the SLR a 5-step process is developed based on systematic review procedures by Kitchenham [Kit04], consisting of (1) question definition, (2) study selection, (3) study relevance and quality assessment, (4) data extraction, (5) data synthesis.

2.1.1 Question definition

To achieve the set objective, the following questions were defined and established:

- **Q1: OGD portals of which countries or regions were analyzed by the previous research?**
- **Q2: What OGD portal assessment frameworks, guidelines, or feature lists have been used in previous studies?**
- **Q3: What usability features were identified as lacking when evaluating OGD portals by previous research?**
- **Q4: What recommendations or strategies have been proposed to improve the usability and accessibility of OGD portals in previous studies?**

Q1 is intended to assess which countries or regions have the most researched open data portals. It also aims to establish whether any countries or regions are doing significantly better or worse compared to others to understand which countries or regions are underrepresented. Q2 is intended to provide an overview of the criteria to consider when evaluating OGD portals. Q2 aims to identify a list of candidate frameworks to be used in developing a composite/integrated framework. Q2, Q3 provides insights of the common weak points of the evaluated portals, while Q3 and Q4 aims to derive proposals

for improving OGD portals, both of which subsequently become part of the developed integrated framework under development.

2.1.2 Study selection

Digital libraries covered by Web of Science and Scopus were used to identify relevant literature. These digital libraries were selected because they provide comprehensive coverage of high-quality peer-reviewed scholarly articles and research papers, essential for rigorous academic research. The search includes articles, conference papers, and book chapters published in English in the last six years, as the topic is dynamic, where new works will cite the most influential prior works. These databases were queried for the keywords *((“open data” OR “open government data”) AND portal) AND (usability OR evaluation OR assessment OR "user-cent*" OR analy* OR quality))*, while the search scope was title, abstract, and keywords. Both of these databases provide sophisticated search and filtering tools, although their capabilities differ. Therefore, two search strings were designed (See Figure 2, 3).

The Scopus advanced search functionality appeared more refined and capable of including exclusion and inclusion criteria within the search string. The search string depicted in Figure 2 only searches the title, abstract, and keywords for search terms that are exact or nearest neighbors. The *W/3* statement permits the proximity of the two word groups to three words. The *W/15* works similarly, meant to find word groups within the length of an average sentence [Cut20]. Variations of the word in American or British English were taken into consideration as well. At this point, 170 results satisfied the search criteria. As the field is dynamic, only publications published within the last six years were chosen (exclusion criteria, EC1). That resulted in 133 publications. Then, every publication in a language other than English was excluded (exclusion criteria, EC2), resulting in 126 publications. Finally, the searches were limited to articles, conference papers, and book chapters (exclusion criteria, EC3). As a result, a total of 108 articles were found in Scopus.

In the Web of Science database, the search was also performed against the title, abstract, and keywords. The main difference is that it searches only for exact nearest neighbors. The query resulted in 405 articles. EC1, 2, and 3 were applied within the platform’s user interface. Applying EC1 resulted in 339 publications, EC2 to those left resulted in 307, and EC3 to those left resulted in 283. Many false-positive outputs were detected, necessitating the application of additional filtering. Publications from the categories of *Computer Science Information Systems, Computer Science Theory Methods, Information Science Library Science, Computer Science Interdisciplinary Applications, Public Administration, Computer Science Software Engineering, and Communication* were selected. As a result, 176 publications were found on the Web of Science platform.

```
TITLE-ABS-KEY (
  ( ( open W/3 data OR open W/3 government W/3 data ) W/15 portal )
  AND (
     usability OR evaluation OR
     assessment OR "user-cent*"
     OR analy* OR quality
  )
)
AND LANGUAGE ( english )
AND ( PUBYEAR > 2017 OR PUBYEAR = 2017 )
AND ( LIMIT-TO ( DOCTYPE , "ar" ) OR LIMIT-TO ( DOCTYPE , "cp" ) OR
    LIMIT-TO ( DOCTYPE , "ch" ) )
```

Figure 2. Search string for Scopus.

```
(TS=( ("open data" OR "open government data") AND portal))
AND TS=
(
usability OR evaluation
OR assessment
OR user-centred OR user-centered
OR analysis OR analyzis
OR quality
)
```

Figure 3. Search string for Web of Science (Additional filtering capabilities used within the UI).

2.1.3 Study relevance and quality assessment

Scopus and Web of Science results were merged together (284 publications). Search results without links to the publication and duplicates were identified and eliminated, resulting in 264 and 236 publications left correspondingly. Papers without access to their full texts were removed (225 left). The title and abstract were then scanned to determine the relevance of the study. The following criteria and corresponding grading were defined:

- the study that focuses on assessing the usability of open data portals receives a grade of 1 (the most relevant);
- the study that mentions the usability assessment of open data portals receives a grade of 2;
- the study that mentions open data portals receives a grade of 3;

- the study that is not associated with open data portals receives a grade of 4.

Consequently, articles that received grades of 3 and 4 were filtered out (85 left). Three articles were excluded due to unclear portal evaluation criteria based on full text screening. Finally, 82 studies remained for further analysis (see Figure 4).

2.1.4 Data extraction

To attain the objective of this SLR, a protocol was developed to organize the process of data collection on each selected study, including (1) descriptive information, (2) information related to study approach and research design, (3) information related to its quality and relevance, and (4) OGD portal assessment-related information. The developed SLR protocol was adapted from [ZCS21, NRC+23] with necessary modifications made to align it with the unique requirements of this research. Descriptive information includes generic metadata and the relevance of the article to this study. The approach and research design part gathers the research questions and study aim. The quality and relevance part notes any additional quality concerns after reading the paper and checking whether the assessment of the OGD portal is the primary purpose of the study. The OGD portal assessment part of protocol collects answers to the literature review questions defined in section 2.1.1. Table 1 provides the above-mentioned SLR protocol. Notably, the protocol does not capture information regarding the weighting system, as the necessity for such a system was identified during a later stage of framework development.

Table 1. Structure of SLR protocol

Category	Metadata	Description
Descriptive information	Article title	The title of the article
	List of authors	The names of the authors of the articles
	Complete reference	The complete source information to refer to the study
	Year of publication	The year in which the study was published
	Type of paper	Journal article / conference paper / book chapter
	DOI / Website	A link to the website where the study can be found
	Number of citations	The number of citations of the article in Scopus, Web of Science
	Keywords	Keywords of the paper as indicated by the authors
	Relevance for this study	What is the relevance level of the article for this study? (high / medium / low)

Approach and research design-related information	Objective / Research questions	The research objective / aim, established research questions
Quality- and relevance-related information	Quality concerns	"Whether there are any quality concerns (e.g., limited information about the research methods used)?"
	Primary research object	Is the assessment of Open Government Data portals a primary research object in the study? (Primary/Secondary/Related to the topic)
OGD portal assessment-related information	Q1: OGD portals of which countries or regions were analyzed by the previous research?	List of countries, which national portals are assessed in the study
	Q2: What OGD portal assessment frameworks, guidelines, or feature lists have been used in previous studies?	OGD portal assessment frameworks/guidelines/feature list applied in the study
	Q3: What usability features were identified as lacking when evaluating OGD portals by previous research?	Stressed out missing usability aspects from OGD portal(s) mentioned in the study
	Q4: What recommendations or strategies have been proposed to improve the usability and accessibility of OGD portals in previous studies?	Recommendations and approaches to increase usability and reach of OGD portals mentioned in the study

2.1.5 Data synthesis

The final step of the SLR is data synthesis. The raw data obtained through the above procedure is systematically analyzed, the main results of which are presented in section 3. As part of this step, portal features related to user usability were extracted from articles identified through the data extraction process as most relevant (grade 1 and 2) to build the framework upon (see Table 2). The extracted features' selected scope complies with the framework's scope. In addition, a list of commonly overlooked usability features

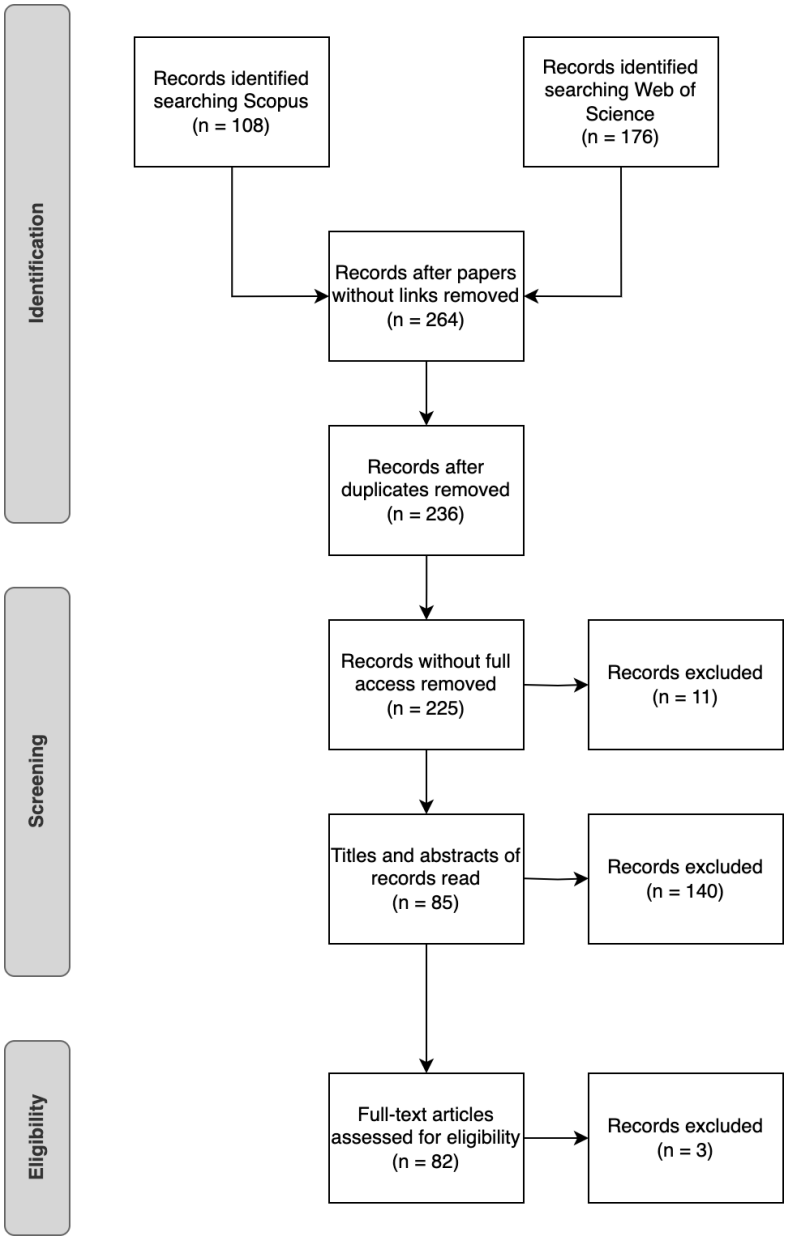


Figure 4. Study selection, relevance and quality assessment (presented using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram).

(Q3, Q4) has been collected and categorized, along with more non-trivial suggestions for improving OGD portal usability beyond refining missed features.

2.2 Prototyping the framework

The process used to design the framework consisted of several steps. By analyzing the data derived for each metadata dimension for the selected articles, "patterns" are derived for sub-dimensions and dimensions of the framework. These are overviewed and prioritized, and then the prototyping of the draft framework starts. In addition to the dimensions and sub-dimensions, the criteria by which the sub-dimensions are deemed satisfactory are specified. The prototype framework, including its aspects (sub-dimensions) and dimensions, was reviewed by 2 experts. The first expert is a master's student who works in the private sector, with over 4 years of experience in SE projects and OGD expertise in research. The second is a PhD holder with over 5 years in academia and practice in SE, HCI, and UX projects, and OGD expertise in research and public administration consultations. This review assessed the sub-dimensions consistency and conformance to the heuristic evaluation method [NM90].

2.3 Conducting portal assessment to test the prototype

During *Stage 3*, the framework prototype is tested to check whether additional sub-dimensions can be added or more general sub-dimensions can be split into multiple ones. The testing is performed on 4 top-performing (based on the ODM report 2022 [Car22], which was the most recent edition at the moment of the design) national OGD portals: French, Irish, Estonian and Spanish. The choice of top-performing EU national OGD portals is related to the assumption that these portals implement many best practices and provide access to emerging trends. The French portal was chosen because it ranks best in ODM report 2022 (the assessment was performed before the ODM report 2023 was published). The Irish, Estonian, and Spanish portals were chosen due to the author's proficiency in the respective national languages of those nations.

The testing was exploratory; in addition, the structure and rationale of the framework were questioned in the context of a live portal. Upon comparing the selected portals, it became apparent that additional sub-dimensions, modifications to the sub-dimension criteria, or the subdivision of a single sub-dimension into multiple ones were necessary. Every time a portal underwent testing, the modifications introduced by its evaluation

Table 2. Studies used for creating the framework

Relevancy	References
High (grade 1)	[LN21], [MHL18], [Kni20], [Car22], [SZCH22], [ALM+18], [AODucDPh22]
Medium (grade 2)	[Nik21], [SGLPM18], [ML17], [DLC18], [WRC18], [GC17], [HMGS20], [FDK20], [RVCKP21], [BBV22], [WRC18], [ZF19], [SVM22], [ADAD17]

were documented, and the affected sub-dimensions of previously evaluated portals were retested. The prototype testing finished when no additional modifications were introduced, and the portals' performances were re-evaluated.

2.4 Compiling the framework

During *Stage 4*, all information sources (SLR, selected articles of experts in portal design, notes from framework prototype testing) and all suggestions were taken into account, resulting in the development of the final version of the framework, which was reviewed again as described in *Stage 2*. It became evident that due to their diversity in terms of nature and importance, these dimensions and sub-dimensions required a weighing system. After the framework was finalized, the weighing system for the score calculation was developed, taking into account the weighing systems encountered in selected publications from portal assessment framework design experts [SZJG15, ZPS21, MHL18, ALM+18, ADAD17, Car22].

2.5 Framework testing and result analysis

During *Stage 5*, the developed framework was tested on a sample of 33 national OGD portals with 27 EU portals and 6 GCC portals. A systematic web search for portals was used to identify relevant portals for each country. The country name was utilized in conjunction with the English search terms *"open data"* and *"open data portal"* on the Google search engine. A resource that maintains an exhaustive inventory of open data portals [HR23] was consulted when the official status of a particular portal is uncertain or inconclusive. Alternatively, the EU national OGD portal addresses can be obtained by consulting the ODM report. The web addresses of the portals are provided in Table 3.

As a result of portal testing:

- each individual portal is assessed, the scores for sub-dimensions, dimensions, and the total score are calculated using the weighing system;
- average scores are calculated for the EU and GCC;
- portals are ranked;
- the top portals (best performers) are determined for each dimension, deriving the best practices and weak points observed on national portals, which is part of the qualitative analysis carried out alongside the quantitative analysis;
- trends in the design of portals are determined and collaborative initiatives between portals are identified.

Table 3. Portals web addresses

Country	Web address
Austria	www.data.gv.at
Bahrain	www.data.gov.bh/pages/homepage/
Belgium	data.gov.be/en
Bulgaria	data.egov.bg/
Croatia	data.gov.hr/en
Cyprus	www.data.gov.cy/?language=en
Czechia	data.gov.cz/english/
Denmark	www.opendata.dk/
Estonia	avaandmed.eesti.ee/
Finland	www.avoindata.fi/en
France	data.gouv.fr
Germany	www.govdata.de/
Greece	www.data.gov.gr/
Hungary	kozadatportal.hu/
Ireland	data.gov.ie
Italy	www.dati.gov.it/
Kuwait	e.gov.kw/sites/kgoenglish/Pages/ OtherTopics/OpenData.aspx
Latvia	data.gov.lv/eng
Lithuania	data.gov.lt/
Luxembourg	data.public.lu/en/
Malta	data.gov.mt/
Netherlands	data.overheid.nl/en
Oman	data.gov.om/
Poland	dane.gov.pl/en
Portugal	dados.gov.pt/en/
Qatar	www.data.gov.qa/pages/home/
Romania	data.gov.ro/
Saudi Arabia	od.data.gov.sa/en
Slovakia	data.gov.sk/en
Slovenia	podatki.gov.si/#
Spain	datos.gob.es/
Sweden	www.dataportal.se/en
United Arab Emirates	bayanat.ae/

An assessment to identify the presence of open-sourced code is a component of the portal testing process. If the portal doesn't provide a link, a web search with the Google search engine is made with a query consisting of the portal's domain name and the phrase "source code" or "GitHub" in English.

To enable a deeper understanding of the relationships and patterns among different portals based on their performance metrics, clustering analysis is employed. Cluster analysis, a fundamental technique in data mining and machine learning, aims to organize a dataset into groups (clusters) based on similarities between data points, allowing for identifying patterns within the data [GS20]. By clustering portals with similar scores, groups exhibiting similar behavior or functionality can be identified, thus providing insights into the overall landscape of OGD portals. Based on the score matrix, two types of clustering analysis (K-means clustering and hierarchical clustering) are carried out, which group similar portals together based on their sub-dimension performance. K-means clustering partitions data into a predefined number of clusters by minimizing the within-cluster variance, while hierarchical clustering constructs a tree-like hierarchy of clusters by iteratively merging clusters based on their similarity [GS20]. Generating the K-Means clusters requires determining beforehand the optimal number of clusters (the Elbow method was chosen due to its wide acceptance and straightforward visual analysis). Applying the Elbow method, a single candidate is selected when multiple candidates exist for the optimal number. Optimal boundary differentiation and keeping the number of clusters reasonably small should guide the decision-making process. On the other hand, generating hierarchical clusters requires building the linkage matrix (the Ward method was chosen due to its wide usage and visual interpretability) and selecting the distance threshold by analyzing the dendrogram of the linkage matrix. In the case of the hierarchical type, such distance threshold must be chosen so that both clustering types generate an equivalent number of clusters.

By calculating the average dimensional scores of portals from both types of clusters, their performance across multiple dimensions is evaluated. The corresponding cluster of another type is chosen based on the greatest number of common portals.

3 Results of the literature review

2021 witnessed the highest volume of articles published, most of which discussed significant pushpower to make data accessible to address common global challenges such as COVID-19 [KH21, HLR+21, Nik21]. This supported the assumption that there is a need to (1) create an integrated framework compliant with the recent trends, (2) re-evaluate the portals to trigger changes according to their pain points, (3) conduct a cross-border and cross-regional evaluation [Com, Com23].

3.1 Regions and countries covered (Q1)

Forty-seven (47) studies examined portals in the European Union (e.g., [LNL+22, ALM+18]). Asia ranks second with twenty-five (25) studies (e.g., [MAL+21, WRC18]), followed by North America with twenty-four (24) studies (e.g., [MAL+21, ZTMA21]). South America received the fewest mentions (13) (e.g., [BBV22, ML17]), followed by Africa (14) (e.g., [ADAD17, ML17]) and the Pacific (14) (e.g., [MHL18, FDK20]). It is important to mention that these regions' government portals are part of a larger sample, meaning that it is uncommon for a study to cover portals from a single region. Research suggests that the most potent OGD portals represent countries of Europe, North America, East Asia, and Australasia [NM21, MAL+21, MHL18].

3.2 Existing frameworks (Q2)

The literature presents a wide range of frameworks, some of which are conceptual and high-level [KRN+18], while others are very detailed and straightforward to use [ML17, LN21].

The Open Data Maturity (ODM) reports, published annually by the European Commission (e.g., [Car22, Kni20]), are a series of reports demonstrating how frameworks adapt to new trends, adding and removing criteria as they progress. Compared to the 2020 report, the 2022 report adds documentation, Application Programming Interface (API) availability, and high-value dataset (HVD) promotions. Regrettably, certain important elements were omitted from the 2022 report, including defining a sustainability strategy, providing source code, conducting performance index dashboards, and satisfaction surveys.

The frameworks are conceptually similar, but they are difficult to compare, which is confirmed by ([SZJG15, AZS21, MLN22, Kao23, LNL+24b]). For this reason, scholars have begun to reuse frameworks that have been shown to be robust, fit their own purposes, or are easy to apply to their own cases [MAL+21].

The usability framework by Machova et al. [ML17] is found the most frequently reused framework (see Table 4). Being among one of the oldest on the list (7 years old), it neglects significant new trends. The transparency-by-design framework [LN21], which

Table 4. Framework reuses (captures the original version of the frameworks; revised versions are not included)

Reused frameworks	Reference	Reuse references
Machova et al. usability framework	[ML17]	[NL21, NM21, Nik20a, MAL+21, Nik20b, Sax19]
"Transparency-by-design"-driven OGD portal assessment framework	[LN21]	[LNL+22]
Information and system quality framework for Greek OGD sources assessment	[ALM+18]	[KAKD22]
African OGD portals' media practitioners' preference-centered framework	[ADAD17]	[Sax18a, Sax18b]

is a revisited version of [ML17], covers more aspects and the latest trends, making it more suitable for adopting. However, the framework maintains its flaws by focusing on transparency rather than usability, especially considering current trends in the resilience and sustainability aspects of OGD portals. For instance, it could benefit from incorporating checks for gamification elements [SZCH22], sustainability-related aspects [LNL+24b], personalization, or more granular multilingualism features.

Out of a total of 21 studies analysed, 4 studies presented a framework that was later used in other studies, namely (1) the usability framework [ML17] reused in 7 studies, (2) "transparency-by-design"-driven OGD portal assessment framework [LN21] and (3) African OGD portals' media practitioners' preference-centered framework [ADAD17] both reused by 2 studies, and (3) "information and system quality framework for Greek OGD sources assessment" [ALM+18] reused by 1 study. While they serve as input for the integrated framework under development, none can be used to assess the state of usability of modern OGD portals granularly.

3.3 Weaknesses of portals (Q3)

As part of Q3, the notorious missing aspects of OGD portals were analyzed and synthesized, grouping them into categories/dimensions.

The most commonly mentioned missing **general portal features** are *poor portal navigation* [MAL+21, SGLPM18, GC17], *lack of prioritization when displaying information* (overloaded with information pages) [MAL+21, FDK20, WIK+17], *multilingualism* [NM21, ALM+18, ES23], and *accessibility features* (e.g., support for screen readers for disabled people) [SGLPM18, FDK20, SMKV23]. In the realm of **data qual-**

ity, challenges such as *metadata absence* [Nik20a, MAL+21, WS20], *inconsistency* [HMGS20], and *data versioning* [Kni20] emerge as significant concerns. The excessive utilization of brief catalog search queries by a user suggests a *lack of confidence* in the search functionality's ability to *deliver relevant data for more extensive queries* [KKI+17]. The issue that unprocessed data cannot be downloaded from the portal [Kni20, ALM+18] continues to be a concern. Voiced concerns include the *scarcity of valuable* or *generally low value* of datasets [Nik21, NL21] and the *lack of data visualization and analytical tools* [DLC18, NM21]. The **feedback and support mechanisms**, including *comment sections and forums*, are either absent, overly general, or of substandard quality [ALM+18, NM21]. As for **portal sustainability**, there is *lack of a strategy* [Reg20, LNL+24b], *lack of performance dashboards* [QDM19]. The identified weaknesses are considered to be checked as part of the framework under development.

3.4 Existing recommendations (Q4)

In general, recommendations suggest implementing missing features or fixing found issues, although more sophisticated ones are also found. For instance, [Nik20a] recommends *limiting the number of free-form fields for metadata and providing predefined options* to address the lack of metadata and inconsistency. [MHL18] suggest keeping dataset descriptions short and concise, and *keeping the amount of registered-user-only actions to the minimum*. Some studies stress the need to *use the Data Catalogue Application Profile for data portals (DCAT-AP) vocabulary to make metadata discoverable and understandable worldwide* [Kli19, ZF19], which at the same time will improve standardization and interoperability.

Some suggest migrating portal systems to more advanced technological platforms, such as *CKAN, DKAN, Socrata* data management systems [ML17, ZF19].

Some studies emphasize the need to *understand users needs and demands* [WS20, LNL+24b] and *attract a wider audience* [NM21]. The study [SM18] suggested *reducing information pollution on a dataset page , including using algorithms to augment/ enhance metadata management*. To encourage users to use longer search queries, the portal are recommended to use *a query recommendation system* and *automatically fill in missing dataset descriptions* [KKI+17, KKI+19, NL21]. [CSO+20, ZJSS22] show that *focusing on lay citizens* and allowing them to conduct *searches in their preferred language (multilingualism)* are both beneficial and critical.

Last but not least, introducing *gamification elements* and using *storytelling to vulgarize content* is expected to increase the attractiveness and understandability of data portals [SZCH22]. The identified recommendations are integrated or considered for integration into the framework under development.

4 Proposed integrated usability framework for evaluating OGD portal usability

The proposed integrated usability framework focuses on: (1) inclusivity, ensuring the portal is accessible to a wide range of users, including both local/internal and external users of different nationalities and countries being available in different languages; (2) supporting and facilitating user collaboration and active involvement/ participation; and (3) facilitating exploration and understanding of data.

The framework (see Table 5-6) consists of 9 dimensions, which are divided into 72 sub-dimensions. The **"Multilingualism"** dimension consists of *4* sub-dimensions related to language support, including interface availability, content availability, and search functionality across different languages. The **"Navigation"** dimension consists of *3* sub-dimensions related to user interface design elements facilitating ease of navigation within the system, including menu structures, breadcrumb usage, and tabs for content-rich pages. The **"General performance"** dimension consists of *4* sub-dimensions related to the overall efficiency and effectiveness of the portal, including load time, responsive design, absence of errors, and ensuring an adequate level of accessibility. The **"Data understandability"** dimension consists of *11* sub-dimensions, encompassing various aspects such as the promotion of high-value datasets, dataset views and downloads, showcasing dataset re-use, providing data previews, visualization tools, and vulgarized content to enhance comprehension through examples and visual aids. The **"Data quality"** dimension consists of *9* sub-dimensions, covering aspects such as the availability of machine-readable data formats, basic metadata elements, update frequency accuracy, temporal and spatial coverage, dataset quality ratings and explanations, as well as the presence of an automated dataset quality checklist. The **"Data findability"** dimension consists of *15* sub-dimensions, encompassing various aspects such as discoverability by the publisher, categories, formats, tags, and license, sorting options, dataset metadata, API and SPARQL endpoints, as well as features like recommender systems and featured topics. The **"Public engagement"** dimension consists of *13* sub-dimensions, including features such as use-case uploads, community-sourced content, social media integration, notification systems, up-to-date information, promotion of events, personalization options, request forms, and request tracking, as well as gamification-related ones such as badges, rewards, quizzes, competitions. The **"Feedback mechanisms and service quality"** dimension consists of *7* sub-dimensions, encompassing features such as portal-wide comment sections or forums, direct communication between publishers and users, dataset-specific feedback mechanisms, usefulness assessment, provision of guidelines and tutorials, support contact options, and suggestion forms for improvement. Finally, the **"Portal sustainability and collaboration"** dimension consists of *6* sub-dimensions, including features such as availability of sustainability strategies, performance index dashboards, mentions of collaboration with regional and international governments, user

satisfaction surveys, and the use of open-source code.

To check the presence of the aspect under consideration, a boolean assessment (1/0) is predominantly used, allowing for the inclusion of additional notes that will contribute to the qualitative analysis. However, in the case of accessibility (c4), the web-based accessibility checker [Acc23] is used, where the portal receives a score of 1 if the tool scores it as 71% or higher (compliance without critical issues).

Sixteen (16) sub-dimensions/aspects *(d2-5,7-8,10-11, e1-3,7, and f10-11 in Table 5-6)* are evaluated on a sample basis with a threshold of 70 percent (10 out of 14 datasets) to achieve 1 point. The grading explanation of the aspect *e4* (dataset update frequency accuracy) is closely related to the grading of *e3* (dataset update frequency): it checks whether at least 70% of the datasets' update frequencies are correct. For example, if the update frequency parameter value is "monthly", the latest modification date should be the current or the previous month. Any modification date will fit as long as the frequency value is "unknown" or "irregular". If the update frequency is specified, but there is no way to verify it (including automatic check indicators), it is not considered fulfilled. For the sub-dimension assessment based on a dataset sample, the dataset sample is created as follows: if the portal supports the option to sort datasets by relevance (popularity) and modification date, the first four (4) and last three (3) datasets from data catalog list constitute a sample. If only sorting by modification date is implemented, the first eight (8) and last six (6) datasets form a sample. If sorting is not implemented - the first eight (8) and last six (6) datasets constitute a sample.

Despite our best efforts to classify the aspects according to their primary dimension, certain elements may find themselves in different dimensions or even serve a substantial purpose in other dimensions as a side effect. In addition, aspects vary in the level of importance of the central ideas of the framework. Therefore, there is a need for a weighing system. To introduce a weighting system, we first consult the literature to identify potential systems that we could reuse or adopt (see Table 7).

Among the most popular approaches is to use equal weights (1) for dimensions and aspects ([LN21, ZF19, ML17], [SZJG15]), (2) for dimensions/aspects ([SZJG15, ZPS21]), (3) equal for dimensions but different for sub-dimensions (aspects) within these dimensions ([SGLPM18, RVCKP21], [SZJG15]). Despite the simplicity of the first option, the approach is often criticized even by those who use it (e.g., [LNL$^+$22]).

Instead, a priority-based option ([SZJG15]) is used, where the score of each aspect is multiplied by its importance in relation to the central concepts of the framework. We define three levels of importance: *low, medium, and high*, which are mapped to *1, 2, and 3* respectively. The overall portal score is determined by adding the multiplied values (1):

$$\gamma = \sum(x_l) * 1 + \sum(x_m) * 2 + \sum(x_h) * 3 \tag{1}$$

where γ is the overall score, x_l, x_m, x_h - is the score of the aspect marked as low, medium, high importance, respectively.

Table 5. Proposed integrated OGD portal usability framework (1/2)

Dimension	Sub-dimension	Description	Criteria	Weight
(a) Multilingualism	(1) English is one of the supported languages	one of the supported languages is English	0 - no, 1 - yes	1
	(2) portal interface is available in the supported languages	content is at least partially translated, ensuring accessibility and usability for a diverse range of users	0 - no, 1 - yes	2
	(3) portal content is available in the supported languages	blogposts, dataset pages, manuals, tutorials are translated to the portal-supported languages, enhancing accessibility	0 - no, 1 - yes	3
	(4) dataset search can be done in English	users can search for datasets in English (the language of international communication), enhancing accessibility and search functionality	0 - no, 1 - yes	3
(b) Navigation	(1) convenient menubar structure	users can easily navigate the website	0 - no, 1 - yes	3
	(2) breadcrumb usage	users are provided with a navigational trail, displaying the hierarchical path and aiding in understanding a webpage's location within a website	0 - no, 1 - yes	2
	(3) tabs for content-rich pages	content-rich webpages are structured with tabbed navigation elements to facilitate organized presentation and easy access to various sections of information	0 - no, 1 - yes	2
(c) General performance	(1) portal loads in less than 4 seconds	the portal should load in 4 seconds	0 - no, 1 - yes	3
	(2) responsive web design	functions effectively and is accessible on smartphones and tablets, ensuring a user-friendly experience for mobile users	0 - no, 1 - yes	1
	(3) no blocking errors or exceptions	within the basic usage of the portal, no unexpected errors or exceptions were encountered	0 - no, 1 - yes	2
	(4) sufficient accessibility level	the accessibility testing is assessed by accessibility checker service [Acc23]. European Accessibility Act (EAA) compliance was chosen to test against. The threshold to be compliant is a score of 85	0 - the score is under 61 (less than 71% of partial compliance without critical issues), 1 - the score is above or equal to 61	2
(d) Data understandability	(1) HVD promotion	the promotion of HVD (High-Value Datasets) in portals involves highlighting datasets that are particularly valuable, relevant, or significant for users and encouraging their access and utilization	0 - no, 1 - yes	3
	(2) dataset views	the number of times a dataset has been accessed and viewed by users on a data portal or platform	0 - no, 1 - yes (10/14)	2
	(3) dataset downloads	the number of times a dataset has been downloaded by users on a data portal or platform	0 - no, 1 - yes (10/14)	2
	(4) dataset re-use/showcase count	the number of instances where a dataset has been utilized or repurposed by users on a data portal or platform for various applications or analyses	0 - no, 1 - yes (10/14)	1
	(5) re-use/showcase display in dataset page	presenting or showcasing instances where data from a dataset has been utilized or repurposed on the dataset's page	0 - no, 1 - yes (10/14)	2
	(6) re-use page dataset list	re-uses are supplied with the list of used datasets	0 - no, 1 - yes	3
	(7) data preview	preview of a dataset, offering a sample or snapshot of its content to provide users with a quick understanding of its structure, format, and potential value.	0 - no, 1 - yes (10/14)	3
	(8) data visualization, analytics, and filtering tools	software features enable users to visually represent data, analyze it for insights, and refine the information displayed by applying various filters, enhancing data exploration and decision-making	0 - no, 1 - yes (10/14)	3
	(9) interactive data visualization	dynamic and user-engaging graphical representations of data enable users to visually represent data, analyze it for insights, and refine the information displayed by applying various filters, enhancing data exploration and decision-making	0 - no, 1 - yes	2
	(10) data visualization download	the portal enables saving the results of data visualization or data preview	0 - no, 1 - yes (10/14)	2
	(11) vulgarized content(described through examples and visual aid)	simplified or easily understandable description of complex content is provided, making it accessible and relatable to a broader audience without diminishing its quality or value	0 - no, 1 - yes (10/14)	3
(e) Data quality	(1) machine-readable data formats	each dataset is available in machine-readable formats	0 - no, 1 - yes (10/14)	3
	(2) basic metadata elements	dataset is supplied with metadata consisting of at least: title, description, category, publisher, license, modification date	0 - no, 1 - yes (10/14)	3
	(3) update frequency of datasets	the update frequency is specified for datasets	0 - no, 1 - yes (10/14)	3
	(4) dataset update frequency accuracy (actual vs promised)	the dataset update frequency accurately reflects the scheduled frequency at which data within the dataset is refreshed or modified, ensuring users are informed about the dataset's currentness	0 - no, 1 - yes (70%+ of datasets with update frequency)	3
	(5) data temporal coverage	datasets, when appropriate, have the range of time included in them, that is marked in the dataset page	0 - no, 1 - yes	2
	(6) data spatial coverage	datasets, when appropriate, have the geographic area for which this dataset is relevant, that is marked in the dataset page	0 - no, 1 - yes	2
	(7) dataset quality rating	data quality rating of each dataset is provided	0 - no, 1 - yes (10/14)	3
	(8) rating explanation	the criteria (metadata fullness, availability of certain metadata points) for the rating can be found, which provides a deeper insight on a dataset	0 - no, 1 - yes	3
	(9) automated dataset quality checklist	an automated dataset quality checklist brings substantial benefits to portal users by swiftly and accurately assessing dataset quality, ensuring that the data they access is reliable, up-to-date, and of the highest standard, enhancing their overall experience	0 - no, 1 - yes	3
(f) Data findability	(1) discoverability by publisher	datasets can be filtered by publishers	0 - no, 1 - yes	3
	(2) discoverability by categories	datasets can be filtered by categories	0 - no, 1 - yes	3
	(3) discoverability by formats	datasets can be filtered by formats	0 - no, 1 - yes	3
	(4) dataset format list	list of formats available for the dataset are visible from the catalog (e.g. in the form of tags), providing quick information about the data retrieval methods	0 - no, 1 - yes	2
	(5) discoverability by tags	datasets can be filtered by tags	0 - no, 1 - yes	2
	(6) discoverability by license	datasets can be filtered by licenses	0 - no, 1 - yes	1
	(7) sorting by modification date	dataset sorting that enables users to organize and arrange datasets based on modification date, facilitating efficient data discovery and access	0 - no, 1 - yes	3
	(8) sorting by relevance	dataset sorting that enables users to organize and arrange datasets based on relevance, facilitating efficient data discovery and access	0 - no, 1 - yes	3
	(9) sorting by dataset metadata	sorting that enables users to organize and arrange datasets based on metadata criteria, facilitating efficient data discovery and access	0 - no, 1 - yes	2
	(10) dataset tags	datasets have descriptive labels or keywords added to enhance searchability and categorization	0 - no, 1 - yes (10/14)	2
	(11) data download	data is directly accessible for download on the portal, eliminating the need for users to navigate to external sources or websites for data retrieval	0 - no, 1 - yes (10/14)	3
	(12) API endpoints	API endpoints are available	0 - no, 1 - yes	3
	(13) SPARQL endpoints / RDF files	SPARQL endpoint or linked data is available	0 - no, 1 - yes	2
	(14) recommender system	users are provided with preferably personalized suggestions or recommendations for datasets based on their preferences, interests, and prior interactions with data, making it easier for them to discover relevant and valuable information	0 - no, 1 - yes	3
	(15) featured topics	users are offered pages that compile curated data collections related to a specific theme or concept to show a contextual bigger picture	0 - no, 1 - yes	3

Table 6. Proposed integrated OGD portal usability framework (2/2)

Dimension	Sub-dimension	Description	Criteria	Weight
(g) Public engagement	(1) use-case upload feature	users are provided with the opportunity to submit use-cases	0 - no, 1 - yes	3
	(2) community-sourced / citizen-generated data	users are allowed to upload community-sourced or citizen-generated data to the portal	0 - no, 1 - yes	2
	(3) social media support	links to its official pages on popular social media platforms such as Facebook, X (formerly Twitter), LinkedIn, or others are provided. These links facilitate communication, updates, and engagement between the open data portals and their user communities through social media channels	0 - no, 1 - yes	3
	(4) notification system	enables users to subscribe to receive notifications or newsletters, keeping them informed about updates, news, and relevant content pertaining to the portal's activities and offerings	0 - no, 1 - yes	2
	(5) portal up-to-date information	the information on the portal is current and up-to-date, ensuring that users have access to the most recent and relevant content	0 - no, 1 - yes	3
	(6) sessions and events promotion	the portal provides information about meetings, workshops, or gatherings designed to raise awareness, provide training, or engage the public in using and benefiting from the national portal	0 - no, 1 - yes	3
	(7) personalization features	additional features to non-publishers, offering enhanced functionalities and capabilities, are provided. Examples of those features could be: a personalized list of favorite datasets, subscription to topics, comment mentioning system with e-mail notifications, badge collecting	0 - no, 1 - yes	1
	(8) badges	virtual achievements or symbols awarded to users for completing specific tasks or reaching milestones, adding an element of accomplishment and recognition to the portal are provided	0 - no, 1 - yes	3
	(9) rewards	users are provided with tangible or virtual incentives to encourage desired behaviors and participation, community support	0 - no, 1 - yes	1
	(10) quizzes	interactive assessments are provided, engaging users in knowledge testing, promoting learning and user engagement through questions and challenges	0 - no, 1 - yes	3
	(11) competition	users compete against each other to achieve specific goals or rankings, fostering engagement and motivation	0 - no, 1 - yes	2
	(12) request forms	web-based tools that users can utilize to formally request specific datasets from the data providers or the data portal itself are provided	0 - no, 1 - yes	3
	(13) request tracking	page that allows users to monitor the progress and status of their requests, providing them with real-time updates and information about the handling of their inquiries or demands	0 - no, 1 - yes	3
(h) Feedback mechanisms and service quality	(1) portal-wide comment sections or forums	online discussion areas where users can engage in conversations, share information, or express their opinions on various topics or aspects of the portal	0 - no, 1 - yes	3
	(2) direct publisher-user communication	users can engage in direct and immediate interaction with publishers, facilitating feedback, questions, or discussions.	0 - no, 1 - yes	3
	(3) comment sections or forums for datasets	online spaces associated with individual datasets, where users can post comments, questions, and discussions related to that specific dataset, fostering communication and collaboration around the data	0 - no, 1 - yes	3
	(4) dataset usefulness assessment	feature allowing users to mark a dataset as useful, often through actions like upvoting or liking, enhances user engagement and provides a way to highlight valuable datasets	0 - no, 1 - yes	3
	(5) guidelines, tutorials, manuals, FAQs	informative resources designed to help users understand and utilize a service or product effectively, providing guidance, instructions, and answers to common questions.	0 - no, 1 - yes	3
	(6) contact for support	means for users to get in touch with a support team or customer service to seek assistance, ask questions, or report issues related to a product or service	0 - no, 1 - yes	3
	(7) improvement suggestion form	suggestion for improvement form is a tool that allows users to provide feedback, ideas, or recommendations to enhance the portal, fostering user engagement and continuous enhancement of the platform's features and services	0 - no, 1 - yes	3
(i) Portal sustainability and collaboration	(1) sustainability strategy	strategic plan that outlines how the portal aims to ensure the long-term availability, relevance, and impact of the data it hosts, often encompassing funding models, data governance, user engagement, and partnerships to support ongoing operations and growth	0 - no, 1 - yes	2
	(2) performance insights dashboard	visual representations and numerical data that provide a quick and easily digestible overview of key performance indicators, allowing stakeholders to assess the effectiveness and progress of the portal	0 - no, 1 - yes	3
	(3) regional governments collaboration mentions	information about cooperative efforts and partnerships with local governments or institutions	0 - no, 1 - yes	2
	(4) international collaboration mentions	information about cooperative efforts and partnerships with international governments or institutions	0 - no, 1 - yes	3
	(5) user satisfaction survey	assessment to collect feedback and opinions from users regarding their experiences and contentment with the services, data accessibility, and overall performance of the portal is conducted	0 - no, 1 - yes	3
	(6) open source codebase	code is open source, and the link to the repository is publicly available on the portal	0 - no, 1 - yes	1

Table 7. Weighing systems used in the existing research. *Asterisks (*) means multiple indices are mentioned in 1 study*

Category	References
Equal weights for dimensions and aspects	[LN21, ZF19, ML17], WJP Open Government Index [SZJG15]*
Equal weights for dimensions, different for aspects	[SGLPM18, RVCKP21], ODI Barometer, ePSI Scoreboard [SZJG15]*
Different weights for dimensions and aspects	[Car22, Kni20, HMGS20, WCR18], Open Data Watch [ZPS21]*
Equal weights for dimensions	Capgemini OD Economy [SZJG15]*, Open Government Data Report [ZPS21]*
Different weights for dimensions	[ADAD17], OKNF OD Index [ZPS21]*
Importance/priority factor involved	World Bank ODRA [SZJG15]*
Aggregated results	[ALM$^+$18, Nik21, AODucDPh22]

The maximum possible result for the portal assessment is 177 points, of which 9 are obtained in the *Multilingualism (a in Table 5-6)* dimension, 7 - in *Navigation (b)*, 8 - in *General performance (c)*, 26 - in *Data understandability (d)*, 25 - in *Data quality (e)*, 38 - in *Data findability (f)*, 32 - in *Public engagement (g)*, 18 - in *Feedback mechanisms & service quality (h)*, 14 - in *Portal sustainability & collaboration (i)*.

5 Application of the developed framework to 33 OGD portals: analysis of the results

The developed integrated OGD portal usability framework was applied to 33 EU and GCC OGD portals, where the individual portal scores were calculated using Equation 1 for each dimension, which were then summed to produce a total portal score.

France is found to be the leader in the portal ranking, as shown in Figure 5, which is consistent with the results of previous studies (e.g., [Car22]). However, the French portal did not receive the maximum score (177), having some room for improvement. Nevertheless, its score of 141 is objectively high, leaving Saudi Arabia (122) 19 points behind in second place. Those results are highly competitive considering the average portal score is 84.9, the score for EU-only portals is 88.7, and for GCC portals, it is 67.8. Eleven national OGD portals passed the 100-point threshold, and only four scored below 50, with OGD portal of Kuwait coming in last place.

The Kuwaiti OGD portal's low rank is most likely related to the fact that Kuwait does not have a national open data portal. The Kuwaiti government's portal has only a dedicated OGD provision section. Concerning the search of national OGD portals, obstacles appeared not only in the Kuwait case. The Hungarian [KM24][ABAK23] community-sourced portal is the only one appearing in the Google search results (following the Section 2.5 sample selection methodology). The portal contained a disclaimer that explicitly mentioned its community-sourced nature. This prompted a verification of the portal's address in the ODM report, after which an official one was identified.

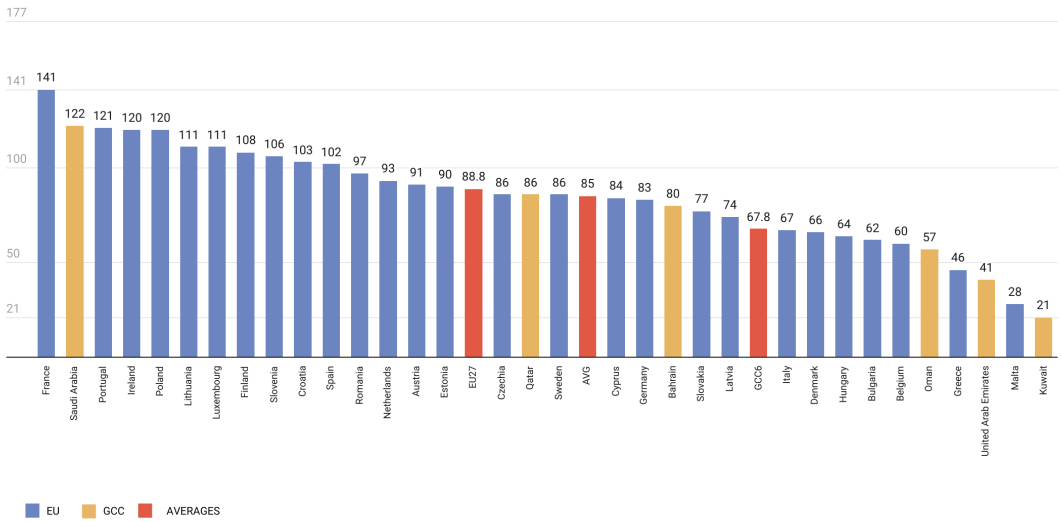


Figure 5. Portal ranking

Let us now discuss the results demonstrated by the selected portals by dimension.

5.1 Multilingualism

In the "Multilingualism" dimension, eight countries' portals received the maximum score (9): **Bahrain, Estonia, Ireland, Malta, Oman, Qatar, Saudi Arabia, and the United Arab Emirates** - as shown in Figure 6. Five of them are GCC states (5 out of 6). Their portals' extensive support for the English language interface illustrates these nations' willingness to collaborate beyond the region. Conversely, the majority of EU portals have a minimal positive score. The portals of *Germany, Hungary, and Italy* have not achieved any points because they exclusively support German, Hungarian, and Italian, respectively. The *Irish and Maltese* portals provide English language support as it is one of their official languages.

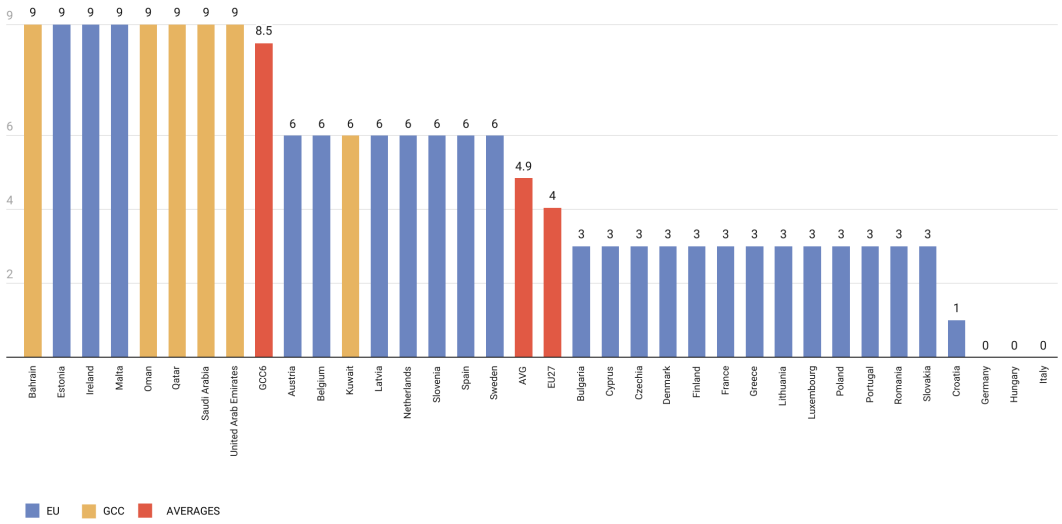
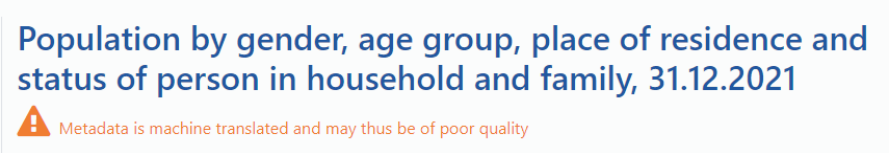
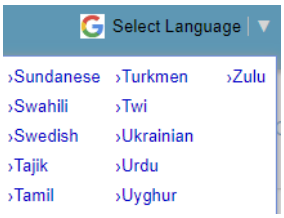


Figure 6. Ranking in "Multilingualism"

However, the *Estonian* portal demonstrates the successful use of machine-translated metadata. The portal translates metadata and makes the users aware of possible low-quality metadata translation (see Figure 7a). The *Austrian* portal only offers to redirect the user to the EU Open Data portal to access machine-translated metadata in the selected language. In addition, publishers can optionally include the title and description in English (the second supported language). The *Slovenian* portal implements portal-wise translation through the Google Translate plug-in, which translates the content in addition to the user interface (see Figure 7b). The disadvantage of this approach is that the search still relies on the original metadata of the dataset, which is mainly in Slovenian.



(a) The Estonian portal machine-translates metadata



(b) The Slovenian portal's Google Translate plug-in

Figure 7. Examples of "Multilingualism" dimension features

Overall, the level of multilingual support is fairly basic among analyzed portals and should be improved, with a particular focus on allowing users to search datasets in their language of choice.

5.2 Navigation

In the "Navigation" dimension, fourteen (out of thirty-three) countries' portals received the maximum score (7): **Austria, Croatia, Denmark, France, Ireland, Latvia, Lithuania, Luxemburg, Poland, Portugal, Romania, Saudi Arabia, Slovakia, Sweden** - as shown in Figure 8. Only *Saudi Arabia* represents the GCC states among the top performers. *The United Arab Emirates (UAE)* portal significantly impacted the regional average by failing to score points.

An element that caught our attention (but was not factored into account in the final score) was the inclusion of *a site map* on the portals of the *Czech Republic, Germany, Poland, and Italy* (see Figure 10). This can serve as an example of good practice that other portals can consider to simplify navigation for users (especially first-time users), as well as to provide an understanding of the portal's functionality and overall landscape.

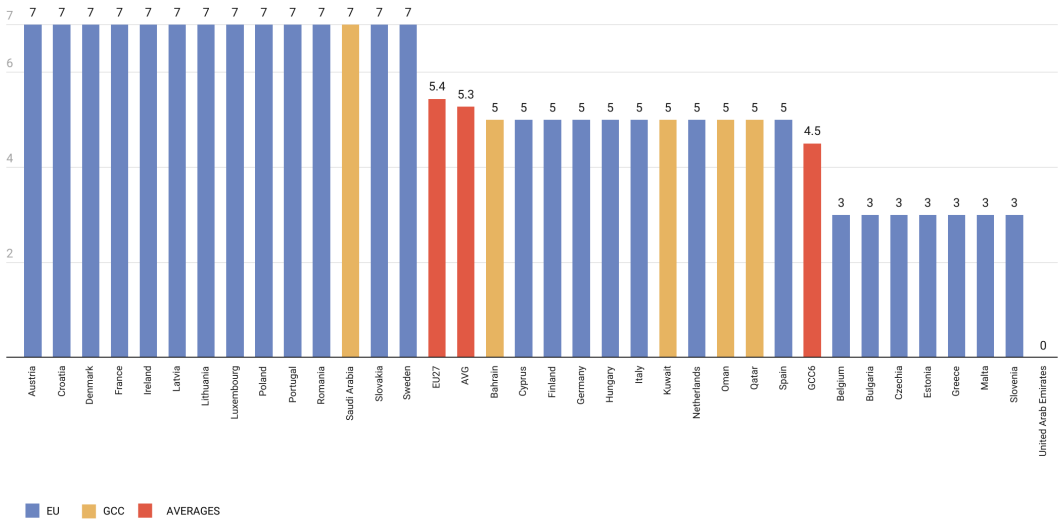


Figure 8. Ranking in "Navigation"

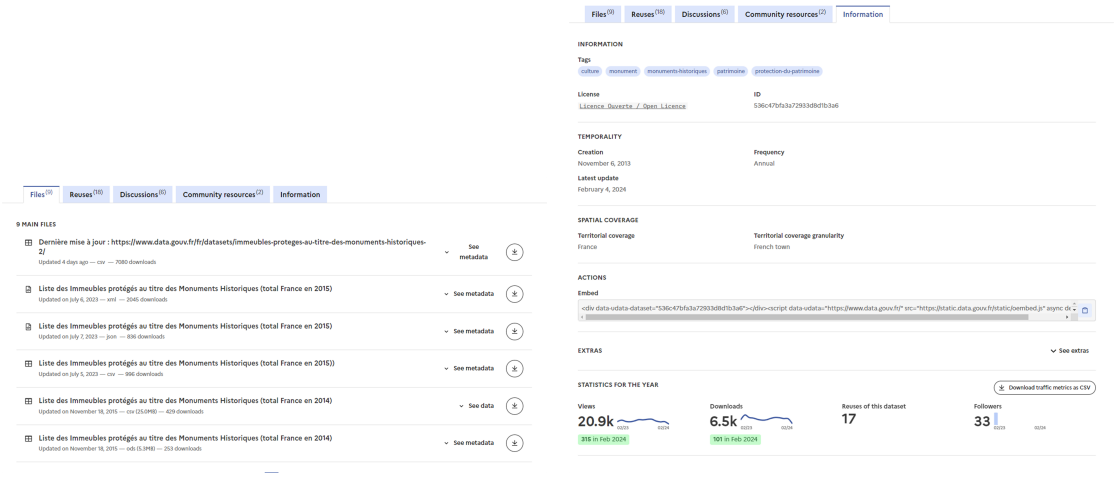


Figure 9. Tabs on the dataset page of the French portal assist in organizing data

As an example of intuitive navigation, all *primary links are placed prominently in the footer* of the *Polish* portal, facilitating effortless access to each section of the portal. *Usage of tabs* filled with a balanced amount of information, as well as intuitively *interconnected pages* can be seen on the *French* portal (see Figure 9).

However, the mere existence of navigation elements does not necessarily make them intuitive or practical. For example, *breadcrumbs* on the *Austrian* portal may not

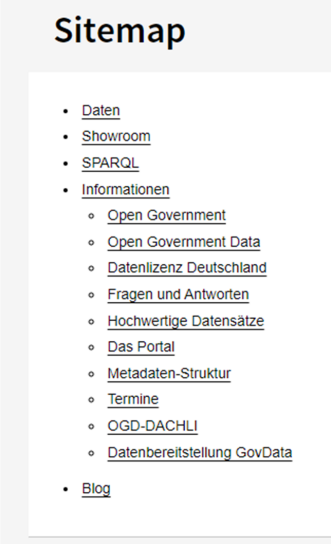


Figure 10. The German portal's sitemap

direct you to previously opened pages. Clicking on the tabs on the dataset page on the *Bulgarian* portal will take you to a new page without a link back to the dataset's main page. The *priority of some portals' tabs* may seem controversial: on the dataset page, the *Portuguese* portal displays publisher information prioritized over the dataset metadata although there are publisher-specific pages for that (see Figure 11). Additionally, the *UAE* portal hides the menu bar on dataset pages, making the content less *consistent* (with the menu bar on pages other than the dataset).

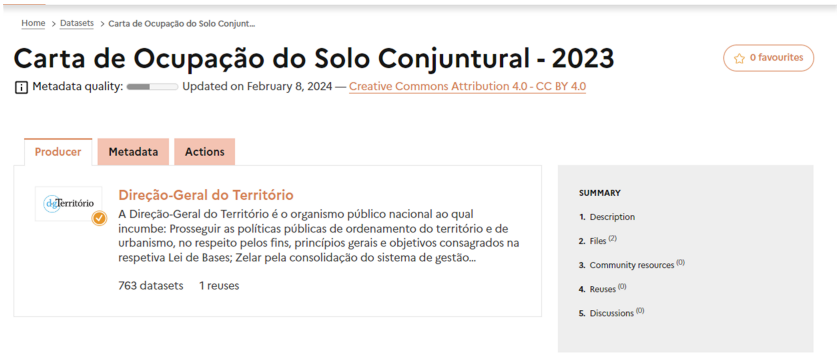


Figure 11. In the dataset page, the producer's information tab is displayed by default instead of metadata information

Portals typically provide essential navigation elements, but some lack consistency.

Additionally, in some cases (e.g., the *Irish* portal), menu bar sections may be flattened or simplified, making a page search easier.

5.3 General performance

In the "General performance" dimension, fifteen (out of thirty-three) countries' portals received the maximum score (8): **Austria, Belgium, Croatia, Finland, France, Italy, Latvia, Lithuania, Luxembourg, Netherlands, Poland, Portugal, Slovenia, Spain, Sweden** - as shown in Figure 12. None of the GCC states has reached the top.

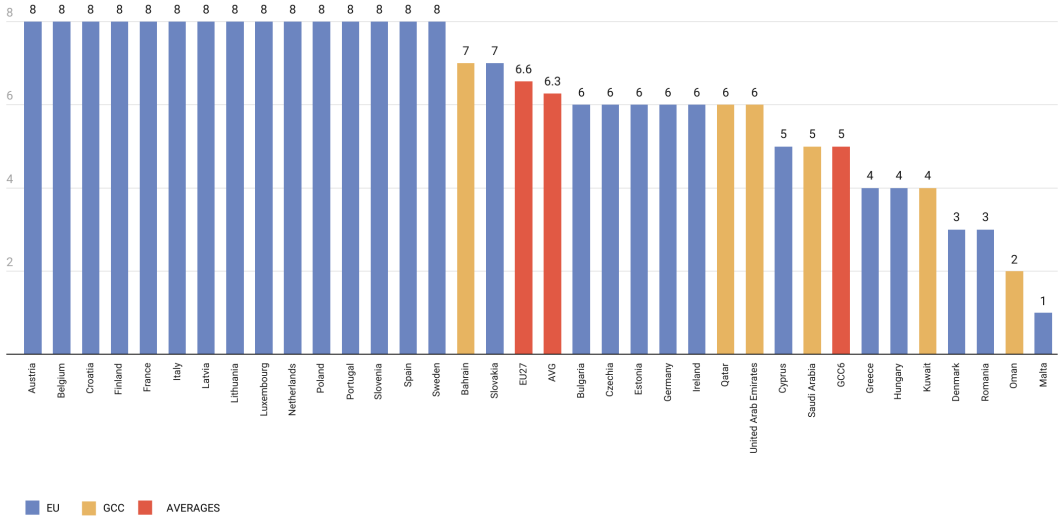


Figure 12. Ranking in "General performance"

For most portals analyzed, the *general performance* is acceptable. The portals of *Poland, Bahrain, Estonia* and *France* serve as exemplary models of responsiveness and robustness.

However, some problems have been identified. In the *Czech* portal, a "delete catalog" (delete dataset) button is present on the dataset page, which is visible not only to the dataset owner (See Figure 13). It is unclear why regular users would have access to the button that leads to the instructions on creating a deletion request for the dataset. It is important to note that this feature is not connected to data error reporting, which sends reports to data curators. The *Maltese* portal displays internal identification keys on its pages. The *Omani* portal has technical restrictions when downloading data: the portal has a limit on the number of rows (2 million) the user can download. The *Irish* portal is not loading some dataset pages (e.g, the last two pages of the oldest / least recent datasets). The *Greek* OGD portal requires registration to access data (API tokens to

download them); however, when attempting to register, no confirmation was received. In addition, the unavailability of dataset resources (files) is common on portals, where only portals of *Latvia, Luxembourg, Portugal, Qatar and Saudi Arabia* had resources for all sampled datasets. The dataset sample, however, may not accurately reflect the overall situation regarding resource availability.

In general, portals could benefit from more emphasis on assessing the quality of portal functionality and improving page loading speed, which currently was observed to be rather slow.

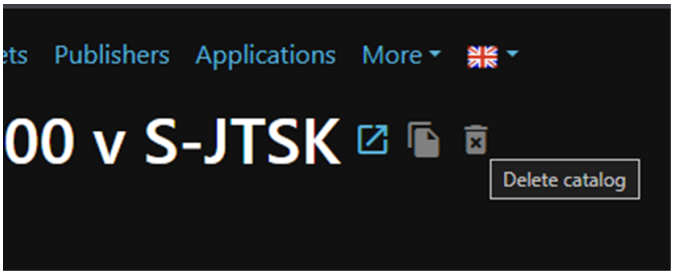


Figure 13. In the Czech portal's dataset page a "delete catalog" button is displayed for any user

5.4 Data understandability

In the "Data understandability" dimension, no portal received the maximum score (26), as shown in Figure 14. The *Polish* portal attains the highest score, while the *Qatari, French, and Saudi Arabian* portals lag slightly behind. Among the top five performers, three are from GCC states. Seven portals: *Bulgarian, Estonian, Italian, Kuwaiti, Maltese, Swedish and UAE* - received a score of zero.

Although it is difficult to determine whether a portal is providing *vulgarized content*, within this dimension, *visualization for datasets*, such as those provided for some datasets on the portals of *Qatar (see Figure 16a), Bahrain, Saudi Arabia, Lithuania, Spain and Poland*, can serve as effective illustrations of making it easier to understand datasets through visual aid.

The *impact section* (e.g., the *Dutch* portal) and *success stories (also use-cases/showcases/re-uses)* (e.g., the *French, Portuguese, German* portals) of applications and services built on open data also serve this purpose. Unfortunately, the latter is the most effective and most resource-consuming.

HVD promotion takes the form of (a) additional filtering criteria on portals of *Ireland, Lithuania, Poland and Slovenia*, (b) promotion on top of the catalog page (*Polish* portal), (c) featured list (*Dutch* portal), or (d) reports highlighting the most valuable datasets

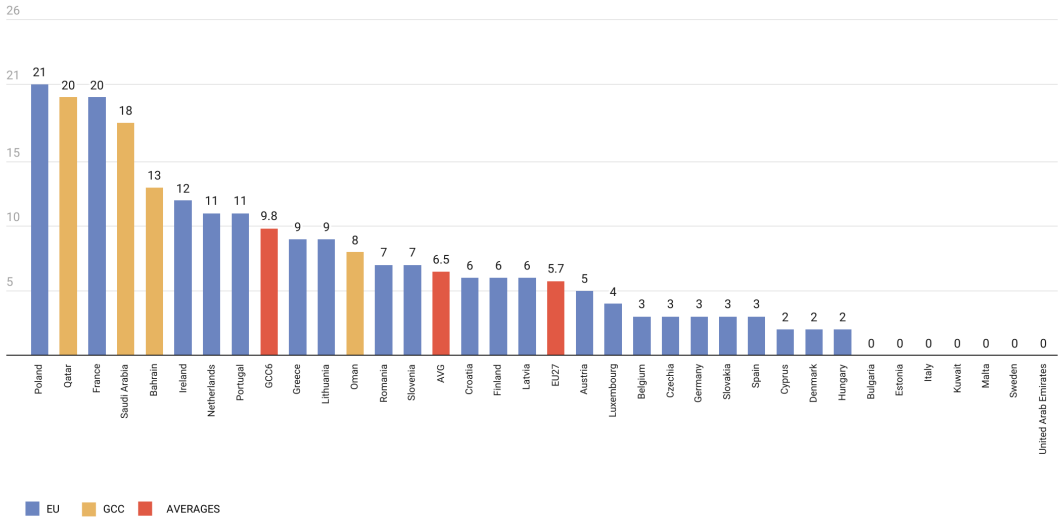


Figure 14. Ranking in "Data understandability"

(*French* portal). The *Czech* portal promotes HVD by holding publishers accountable for the value of their datasets by exposing a variety of indicators (quality of metadata, file, API, schemes, and documentation unavailability) and dashboards to measure dataset performance.

This dimension receives the lowest scores on average. Portals should put more efforts to describe how the data is used (with prior determination of these reuses), what it indicates, and how it can be beneficial. Particular emphasis should be placed on promoting high-value datasets and introducing means to visualize and analyze them. In order to aid users in comprehending the content presented in other sub-views, visual cues should be utilized. For instance, the *French* portal exhibits insightful statistics and charts on the "info" tab of the dataset page 15. Nevertheless, this information may go unnoticed due to the absence of signs that those charts can be found there.

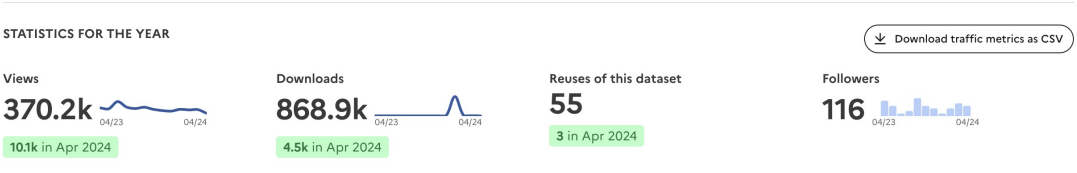
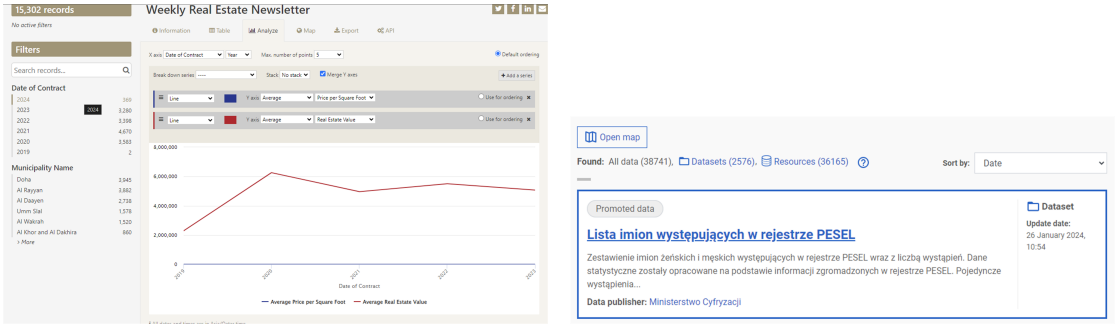


Figure 15. The French portal exhibits insightful statistics and dashboards on the "info" tab of the dataset page



(a) Tool for data analysis and visualization on the Qatari portal

(b) Dataset promotion in the catalog page on the Polish portal

Figure 16. Examples of "Data understandability" features

5.5 Data Quality

In the "Data Quality" dimension, only the **French** portal attained the maximum score (25), as shown in Figure 17. Only the *Saudi Arabian* portal (from GCC) reached the top 10, ranking 9th with the *Slovakian* portal. The *Kuwaiti* portal is the only one that received a score of zero.

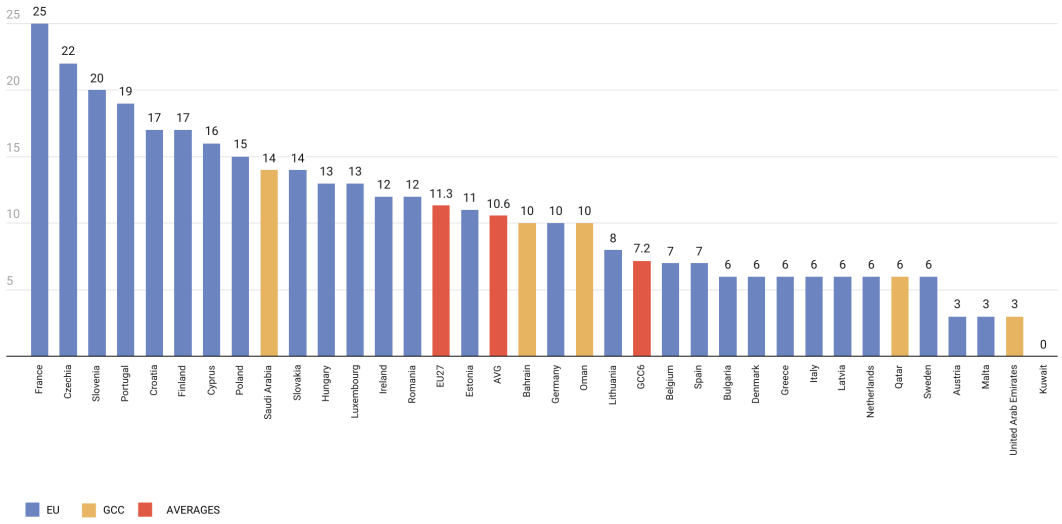


Figure 17. Ranking in "Data Quality"

A common feature of top performers in the *Data quality* dimension - *France, Czechia, Slovenia, Portugal, and Croatia*, is the presence of *indicators*, which provide insights into the availability and accessibility of various aspects of the datasets, including resources,

specifications, and update frequency accuracy. The *Czech* portal additionally checks whether datasets contain personal data (see Figure 19a). The *French and Portuguese* portals share a dataset metadata quality indicator that lets users and publishers know if metadata details are missing or inaccurate (see Figure 18).

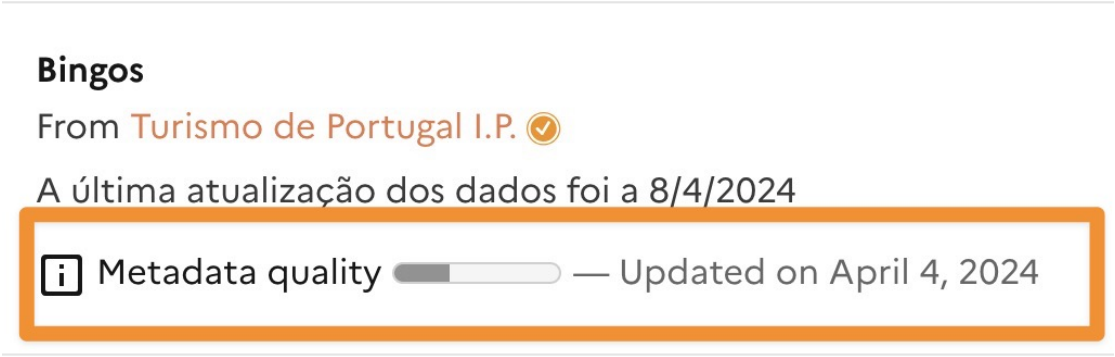
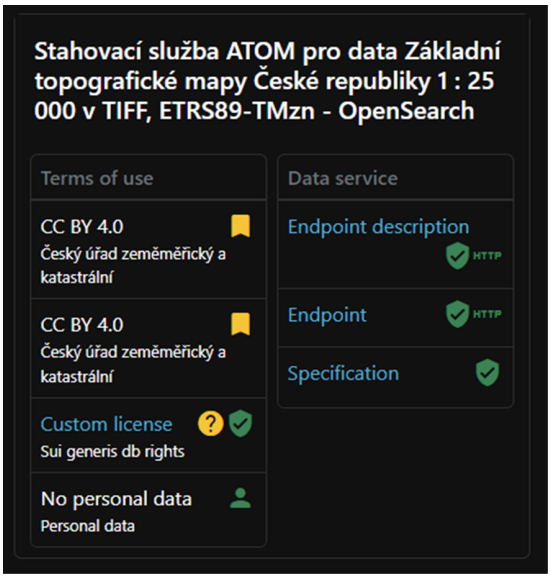
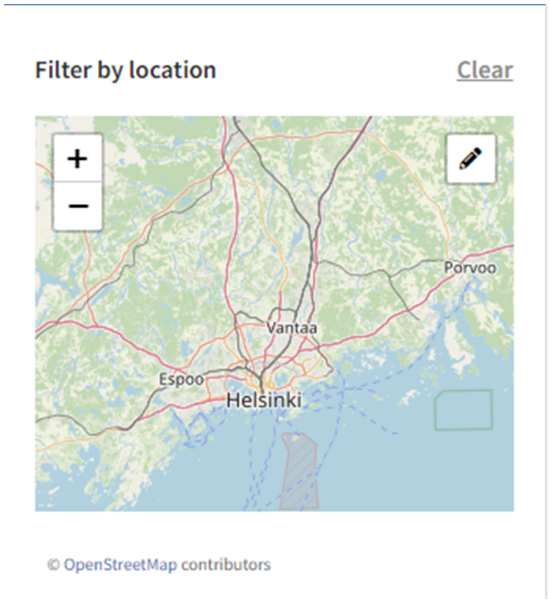


Figure 18. Portugal's portal dataset metadata quality indicator



(a) The Czech portal checks whether datasets contain personal data



(b) The Finnish portal integrates spatial coverage into catalog filtering

Figure 19. Examples of "Data Quality" features

The *Romanian, Slovenian and Irish* portals have *5-star scheme dataset openness rating*, making it easier to determine the openness of datasets.

The *Bahrain and Qatar* portals provides extensive *explanations of dataset schema*. Portal of *Qatar* allows the *catalog to be downloaded in RDF format* (from the DCAT vocabulary). The *Polish* portal allows *dataset metadata to be downloaded in CSV and RDF formats*.

The *Cyprus* portal effectively utilizes the *temporal coverage* parameter. However, the *spatial* parameter may be subject to debate since "Cyprus" value is used for all datasets without more granular division. The *Finnish* portal integrates *spatial coverage into interactive visualizations and even catalog filtering* (see Figure 19b).

The *Romanian* portal has an indicator of resource availability. Nevertheless, that indicator's placement is not prioritized on the resource page (located at the bottom).

In general, portals should improve their metadata collection and provision standards to make it richer and more consistent across datasets. The introduction of various *quality indicators* is a trend that should become widespread. Providing a feature that enables dataset metadata download should be considered an additional benefit.

5.6 Data findability

In the "Data findability" dimension, no portal attained the maximum score (38), as shown in Figure 20. Strong performers (over 30 points) are the portals from *Ireland, Luxembourg, France, Austria and Poland*. The *Saudi Arabian* portal is the only one among the top-10 performers from the GCC states. Only the *Kuwaiti* portal received a score zero.

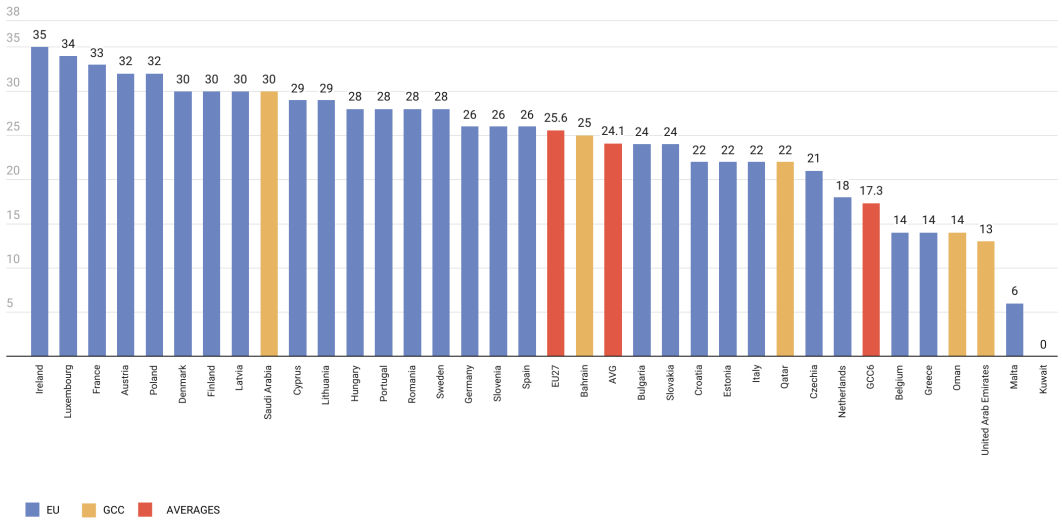


Figure 20. Ranking in "Data findability"

The *Irish* portal provides *advanced search capabilities*, while the *Portuguese* and *Swedish* portals provide user with *search tips* (See Figure 21). The *Finnish* portal provides the ability to *search data within the region selected on the map*.

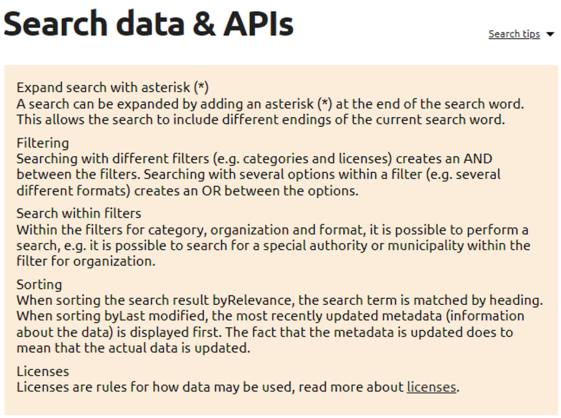


Figure 21. Search tips in the Swedish portal catalog

The *French, Italian, Dutch, Polish, and Luxembourgish* portals provide exemplary illustrations of how to implement *"featured topics"* sections. These topics may be general, but for some portals they are quite specific. For example, the *Polish* portal displays a collection of datasets related to Ukraine, while the *French* portal provides a rich list of topic-specific featured datasets on topics such as energy, education, culture, COVID-19 etc. Related datasets are displayed on the dataset page in the *French* and *Dutch* portals. Unfortunately, the likeness relation/similarity rate is not shown.

The *Omani and Greek* portals demand authentication in order to access their data through the API, which contradicts the openness principle, moreover when it is the only way to download the data in those portals. Similarly, the *Danish* portal exposes endpoints, which return no content.

In general, portals in this dimension perform adequately. However, we suggest they prioritize exposing *API/GraphQL endpoints* and making their content accessible, establishing connections between datasets on similarity to facilitate promotion, and highlighting featured topics.

5.7 Public engagement

In the "Public engagement" dimension, a big gap (11 points) between the best performer (from *Lithuania*) and the maximum score (32) exists, as shown in Figure 22. Among the GCC portals, only the *Saudi Arabian* portal performs better than the average. The *UAE and Malta* portals received a score of zero.

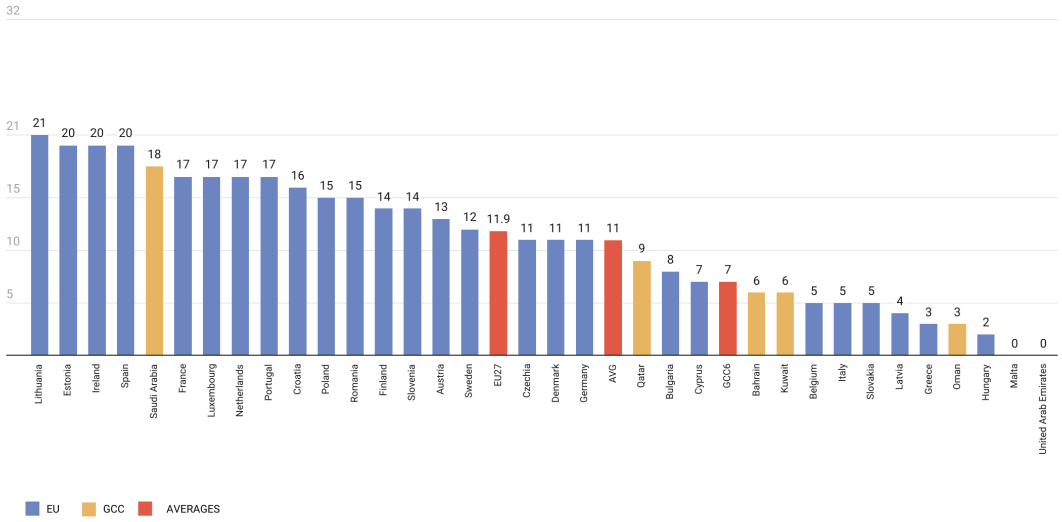


Figure 22. Ranking in "Public engagement"

The *Lithuanian* portal is the top performer in *public engagement* dimension as it has a lively *news section* that has an abundance of articles and event announcements (See Figure 23a). The *Spanish and Croatian* portals have rich *report-tracking* features, displaying a list of reports and the status for each of them (See Figure 23b). In addition, the *Spanish, Estonian, Irish, Lithuanian, Croatian* portals allow users to report a wide range of issues, including data availability and suggestions for improvement, and exchange information regarding reuses and initiatives.

(a) News page in the Lithuanian portal

(b) Request tracking chart in the Spanish portal

Figure 23. Examples of "Public engagement" features

Some portals use video content. For instance, the *Polish, Czech, and Spanish* portals have a decent collection of interviews and educational material posted either directly on

the portal or on YouTube. The *Polish* portal allows users to receive dataset or search result updates via *notifications*. Examples of *community-sourced datasets*, which are maintained not by administrative or governmental institutions, have been seen on portals of *France and Finland*.

On average, the performance of portals in this dimension is relatively low. There is not a single example of introducing advanced gamification elements (e.g., introduction of competition elements, quizzes, rewards, badges). Personalization, in most cases, is limited to following and liking datasets. Many portals lack a catalog of previously reported issues or the possibility of uploading reuses. In addition, social media accounts are rarely active.

5.8 Feedback mechanisms and service quality

In the "Feedback mechanisms and service quality" dimension, *Spain, Croatia, Lithuania, France, Portugal* and *Slovenia* are the top performers in the *feedback mechanisms and service quality* dimension, as shown in Figure 24. However, none of them reached the maximum score (18). Being the only one from the GCC, the *Saudi Arabian* portal once more ranked among the top ten. The portals from *Bahrain, Denmark, Greece, Kuwait and Malta* received a score of zero.

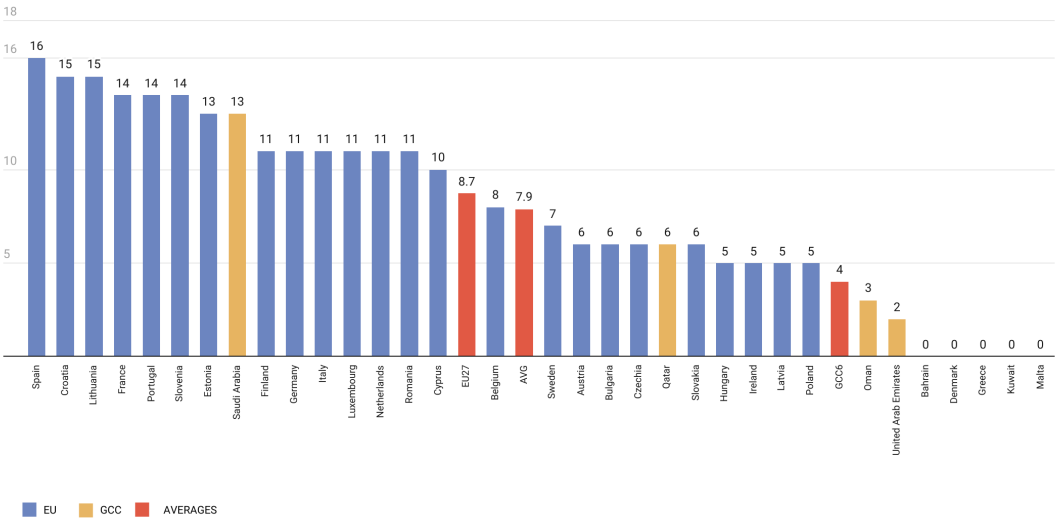
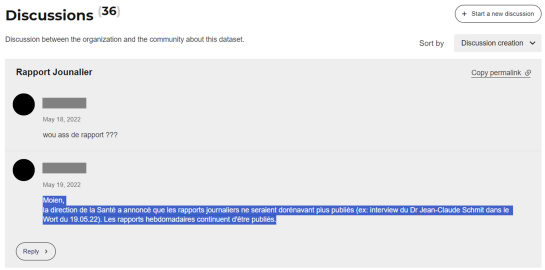


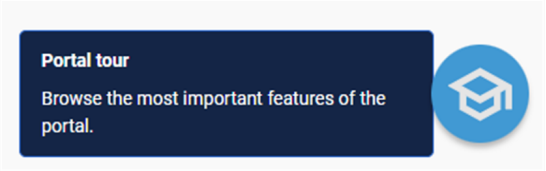
Figure 24. Ranking in "Feedback mechanisms and service quality"

The *comment section* on the dataset page is a trend followed by the portals of *France, Croatia, Lithuania and Luxembourg* (see Figure 25a). The usefulness of this feature can be verified by observing the lively discussion in the corresponding section, where,

however, the participation of both parties, i.e., not only the user but also the publisher, is important.



(a) Lively comment section in the Luxembourgish portal (comments are anonymized)



(b) Polish portal's virtual tour button

Figure 25. Examples of "Public engagement" features

Some portals implement *dataset usefulness assessment* spanning from upvoting/-liking (e.g., the *Croatian, Dutch* ones), 5(10)-score scale (e.g., the *Estonian* portal) to subscribing or following a dataset or page (e.g., the *French, Portuguese, Latvian* ones).

Many portals (e.g., the *French, Portuguese, Austrian, Czech* ones) provide users with - potential publishers and regular users - *guides and manuals*. However, they are often very technical and are unlikely to be understandable to a lay user or tailored to data publishers. Although there are examples (e.g., the *Irish, Dutch, and Czech* portals) where some manuals are tailored toward lay citizens. A virtual tour of the *Polish* portal certain help new users easily navigate the platform (see Figure 25b). The *Saudi Arabia* portal offers a diverse range of *communication channels*: by mail, contact form, address, dataset suggestion or request (the difference between both, however, is not clear since the forms are the same), complaint form.

In general, portals should continue improving their service quality and feedback functions. Portals should create and maintain comment sections to foster communication between publishers and users. The content of manuals and documentation should be more beginner-friendly and rich. There should be forms tailored for different purposes. There are examples (e.g., the *Austrian, Greek, Bulgarian, and Bahraini* portals) where there is no communication with the support service or it only relies on writing emails. Although using online chat for customer service in the *Belgian* portal is uncommon, other portals may find it advantageous to adopt this practice.

5.9 Portal sustainability and collaboration

In the "Portal sustainability and collaboration" dimension, six portals reached the maximum score (14): the **Finnish, French, Irish, Luxembourgish, Polish and Portuguese**

portals - as show in Figure 26. No portals from the GCC states are among the top 10 performers. Only the *Kuwaiti* portal hasn't scored.

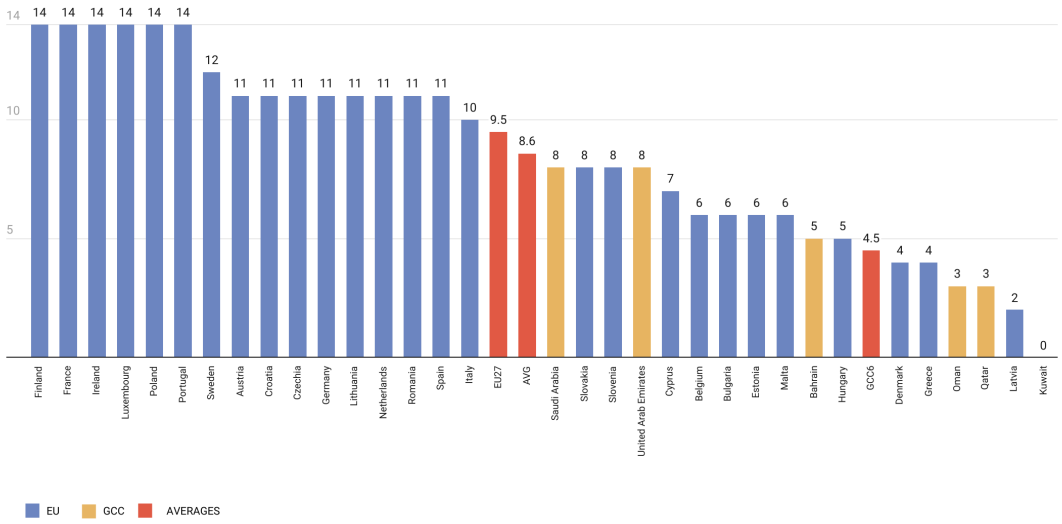


Figure 26. Ranking in "Portal sustainability and collaboration"

Showing a clear sustainability strategy, the *Polish* portal provides the *Open Data Programme for 2021-2027* in the useful material section. The *French* portal *tracks releases of data* every month, highlighting released datasets and reuses (see Figure 27a). The *German* portal celebrated its 10th anniversary in 2023 by sharing past milestones in a blog section.

A comprehensive *user satisfaction survey* is rarely implemented, but the *Finnish and French* portals serve as an example here (See Figure 27b). Instant satisfaction surveys can be found on many pages of the *Saudi Arabian* portal.

The Germany-Austria-Switzerland-Liechtenstein cooperation is highlighted on the portal's strategy pages, demonstrating an example of *international collaboration*, but is not given attention in the catalogs. The *Irish* portal promotes the regional OGD portal of Northern Ireland, while on the *Qatari* portal, the user can build a map based on data connected to the EU states and Israel.

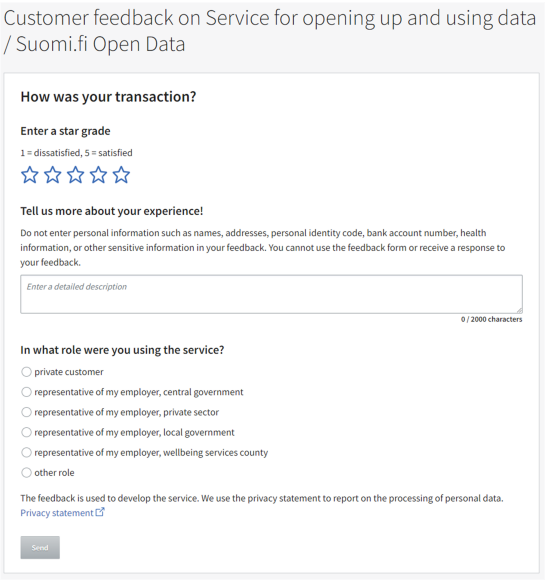
The source code of the portals was found in the associated repositories for 19 of the 33 analyzed portals, but few of the portals post links to the repository on the portal itself. Interestingly, GitHub is the platform of choice for the OGD portals to host their repositories.

On average, portals perform adequately in this dimension. However, we would suggest conducting more user satisfaction surveys, having a defined strategy, sharing reports, and tracking the release of new artifacts (e.g., datasets, visualizations, reuses).

Exit monitoring - January 2024

Published on February 6, 2024

You're reading the January 2024 edition of Release Tracker, an article in which we share the dataset and reuse releases that caught our attention.



(a) French portal release tracker (News section) (b) The Finnish portal user satisfaction survey

Figure 27. Examples of "Portal sustainability and collaboration" features

The benefits of portal partnerships must be highlighted to promote them within and across regions and achieve a more cross-border open data ecosystem.

5.10 Cluster analysis

Clustering analysis facilitates a deeper understanding of the relationships and similarities between various portals characterized by their performance metrics. Two types of clustering analysis (K-means clustering and hierarchical clustering) are performed. They group portals based on their sub-dimension performance, allowing for identifying patterns within the clusters.

Firstly, the clustering groups were produced. The optimal number of clusters was determined to compare the hierarchical and K-means types. The optimal number of K-means clusters was determined using the Elbow method (see Figure 28). The 'elbow' of the graph is the point after which the inertia, or within-cluster sum of squares, starts decreasing at a slower rate [GS20]. For the current case, the 'elbow' is 2; however, partitioning the portals into two clusters will not yield as much insight as if the clusters were more distinct. The secondary 'elbow' (4) was therefore selected. Subsequently, the hierarchical clusters were compiled with the selected number of clusters. It was done by analyzing the dendrogram (see Figure 29) and choosing the appropriate distance

Table 8. Red cluster

Cluster appearance	Countries
Both in K-means and hierarchical clusters	Bahrain, Greece, Kuwait, Malta, Oman, Qatar, United Arab Emirates
Only in K-means	Denmark, Hungary, Latvia

threshold (17) that would produce the desired amount of clusters (4). The code used to generate the clusters of both types can be found in the Appendix.

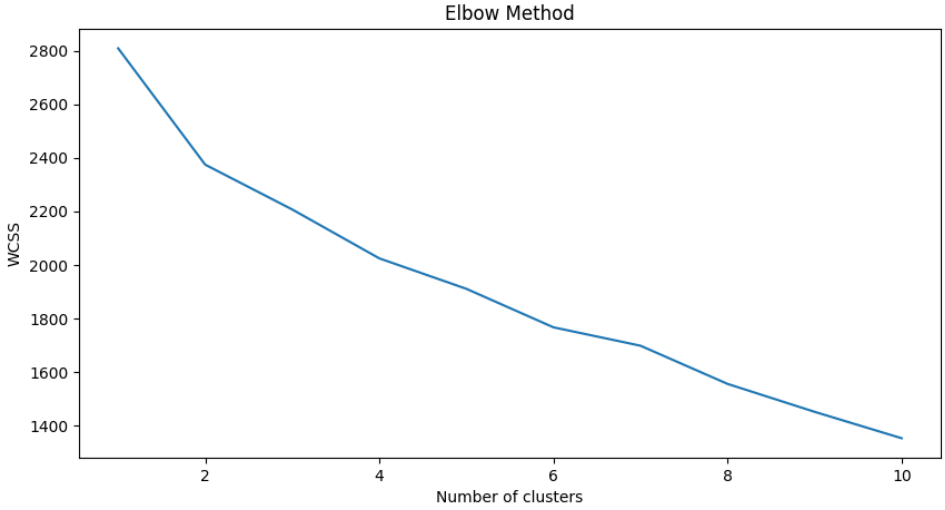


Figure 28. Elbow method (K-Means clustering)

To refer to the merged clusters (from both clustering types), they were assigned color names: red, yellow, blue, and green. Table 8 shows the red cluster, which have the portals of *Bahrain, Greece, Kuwait, Malta, Oman, Qatar, and the United Arab Emirates* appearing in both K-means and hierarchical clusters. The portals of *Denmark, Hungary, and Latvia* appear only in the K-Means cluster.

Table 9 shows the yellow cluster, which consists of the portals of *Belgium, Bulgaria, Cyprus, Germany, Italy, and Slovakia*, appearing both in K-means and hierarchical clusters. The portals of *Denmark, Hungary, and Latvia* (appearing in the K-Means red cluster) currently appear only in the yellow hierarchical cluster. The portal of *Czechia* is the one appearing only in the K-Means cluster.

Table 10 shows the blue cluster, which consists of the portals of *Austria, Croatia, Estonia, Ireland, Lithuania, Netherlands, Romania, Spain, and Sweden*, appearing both in K-means and hierarchical clusters. The portals of *Luxembourg, Poland, and Saudi*

Table 9. Yellow cluster

Cluster appearance	Countries
Both in K-means and hierarchical clusters	Belgium, Bulgaria, Cyprus, Germany, Italy, Slovakia
Only in K-means	Czechia
Only in hierarchical	Denmark, Hungary, Latvia

Table 10. Blue cluster

Cluster appearance	Countries
Both in K-means and hierarchical clusters	Austria, Croatia, Estonia, Ireland, Lithuania, Netherlands, Romania, Spain, Sweden
Only in hierarchical	Luxembourg, Poland, Saudi Arabia

Arabia appear only in the blue hierarchical cluster.

Table 11 shows the green cluster, which contains the portals of *Finland, France, Portugal, and Slovenia*. The portals of *Luxembourg, Poland, and Saudi Arabia* (appearing in the hierarchical blue cluster) appear only in the green K-means cluster. The *Czech* portal (appearing in the K-Means yellow cluster) appears only in the green hierarchical cluster.

The charts in Figure 30 show the merged clusters' performance comparison across nine dimensions.

The red cluster performs the worst (fourth place) in all dimensions, except for *"Multilingualism"* (best score among all clusters) and *"Data understandability"* (3rd place). The average score in *"Multilingualism"* is objectively high (7.01 out of 9), with a substantial gap separating the first and second places (5.39). The areas of difficulty for the cluster are associated with the *"Data quality", "Data findability", "Public engagement", "Feedback mechanisms & service quality", "Portal sustainability & collaboration"* dimensions. Scores from those dimensions are more than twice as low as the maximum score. The portals of this cluster exhibit exemplary approaches relating to the *"Multilingualism" and "Data understandability"* dimensions.

Table 11. Green cluster

Cluster appearance	Countries
Both in K-means and hierarchical clusters	Finland, France, Portugal, Slovenia
Only in K-means	Luxembourg, Poland, Saudi Arabia
Only in hierarchical	Czechia

The third place is occupied by the yellow cluster in numerous dimensions. The yellow cluster performs worst in *"Multilingualism"* and *"Data understandability"*, having scored in the latter just 2.17 out of 26 points. Additionally, the dimensions where the cluster scored less than 50% of the maximum score are *"Data quality", "Public engagement", and "Feedback mechanisms & service quality"*. In other dimensions, the performance is rather acceptable.

The blue cluster in most dimensions is the second best-performing. The blue cluster performed best in *"Navigation" and "Public engagement"*. The areas of difficulty for the cluster are associated with the *"Data understandability", "Data quality"* dimensions. For most dimensions (*"Navigation", "General performance", "Data findability", "Public engagement", "Feedback mechanisms & service quality", "Portal sustainability & collaboration"*), the score gap between the blue and green clusters is around 1-2 points. The portals of this cluster exhibit exemplary approaches relating to the *"Navigation", "Public engagement"* dimension.

The green cluster is the best-performing. It lags behind other clusters only in 3 dimensions: *"Multilingualism", "Navigation", "Public engagement"*. Apart from those dimensions, the green cluster's areas of difficulty are associated with *"Data understandability"*, where it scored less than 50% of the maximum score (as in *"Multilingualism", "Public engagement"*). Although the cluster is the best-performing, only the scores in dimensions *"Navigation", "General performance", and "Portal sustainability & collaboration"* are nearly at their maximum. The portals of this cluster exhibit exemplary approaches relating to the *"Data findability", "Portal sustainability & collaboration"* dimensions.

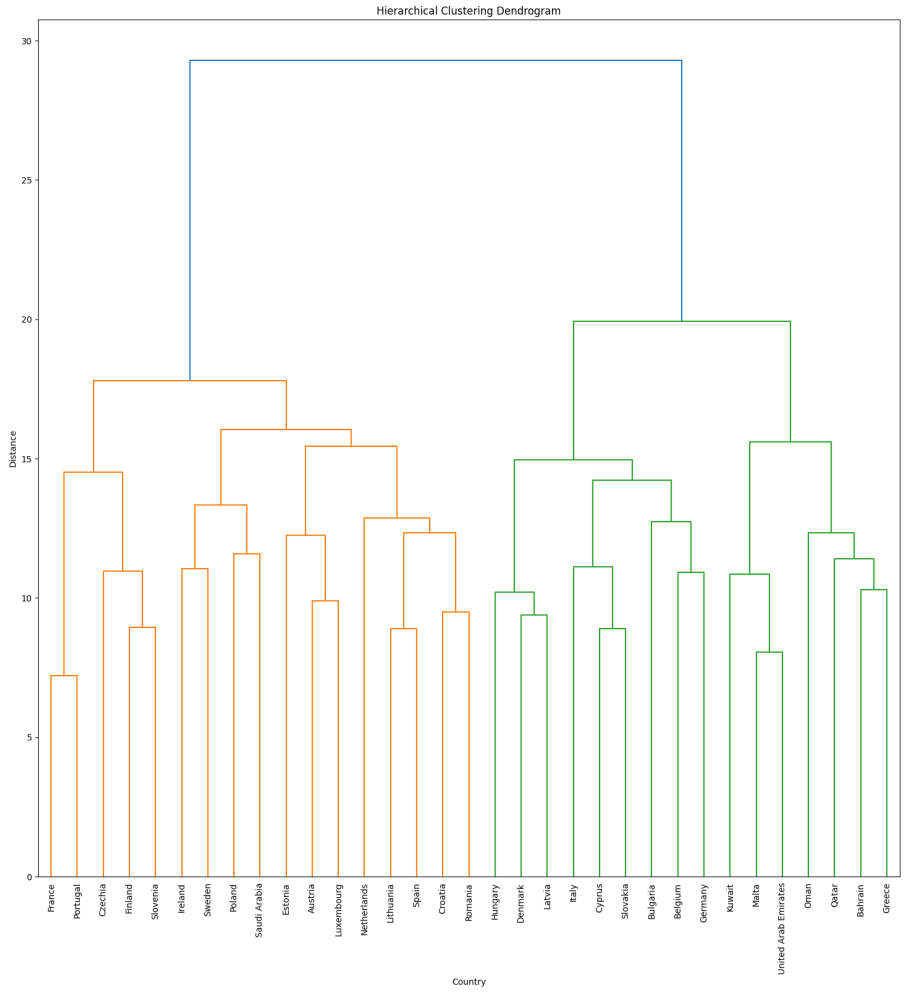


Figure 29. The dendrogram of the linkage matrix for hierarchical clustering

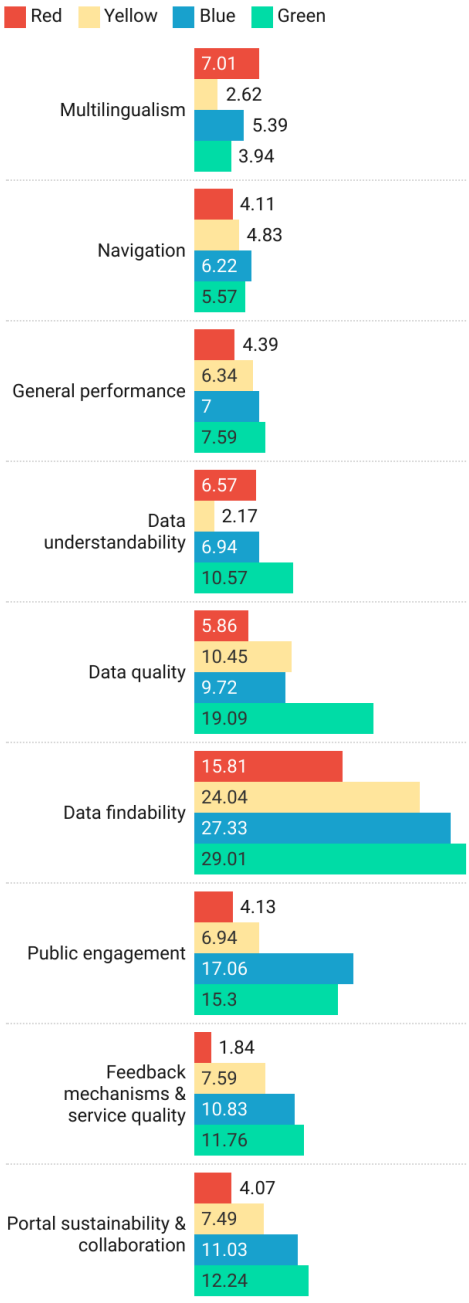


Figure 30. Cluster dimensional comparison

6 Discussion

The integrated OGD portal usability framework provides an instrument for evaluating different aspects and overall performance of the OGD portals regarding (1) inclusivity, ensuring the portal is accessible to a wide range of users; (2) supporting and facilitating user collaboration and active involvement/participation; and (3) facilitating exploration and understanding of data. The observed absence of zero scores for any dimension across 21 out of 33 portals (18 EU, 3 GCC) suggests that portal providers see not only the need to implement the most essential open data portal features (e.g., providing access to non-proprietary machine-readable data) but also features that will attract more users to the portal, improve experience making them more user-friendly, encourage the users to return and discover more data in the portal. Out of those 12 portals that have zero scores in certain dimensions, only 4 have multiple zero scores. That fact implies that portal critical pain points are not scattered across different dimensions, but one mainly. The *"Data understandability"* dimension suffers the most having 7 zero scores, the *"Feedback mechanisms and service quality"* - 5 zero scores, *"Multilingualism"* - 3, *"Public engagement"* - 2. On average, those are the dimensions where portals perform the worst, alongside the *"Data quality"* dimension.

Mapping the current results with those from existing studies is not straightforward because the dimensions/sub-dimensions, weighing system, type of portals under test (city, regional, national portals), and country of origin of these portals differ. For example, comparing performances at local (city and regional) and national levels may be misleading because the performances of open data initiatives at local and national levels may not correspond [LNL+22]. Although differences do not allow a straightforward comparison with the findings of the existing research, some comparison can be still made.

The ODM report series [Pag23, Car22] is probably feature-wise the most up-to-date, although limited to EU ranking. While the dimensions and sub-dimensions differ from those in the framework of the present study, the majority of the portals assessed in this research (27 out of 33) are also assessed in the ODM reports. In the Open Data Portal dimension, the top-5 (only EU) portals in 2023 are of *Poland, Estonia, Ireland, France, Spain* , while in 2022 they were from *France, Poland, Ireland, Slovenia, Cyprus*. The current study's top-5 (only EU) lists portals from *France, Portugal, Ireland, Poland, Lithuania*. Comparing these lists suggests that the current study's framework succeeds in identifying the top performers across various dimensions. On the other hand, the line-up differs when discussing medium and low performers. For instance, *Croatia and Czechia* are among the worst in the ODM reports, however in the current study's result *Czechia* is closer to the EU average and *Croatia* is the 9th best performer among the EU states. This may be explained by more granular sub-dimensions in the integrated framework presented in this study, which may shed more light on aspects that do not factor into the score in the ODM framework. Alternatively, the difference may be explained by the

fact that ODM reports tend to be based at least partially on self-evaluation reports, and there may be an incentive to perform well relative to peers leading to some data collected being unreliable [LNL+24a]. In contrast to the ODM framework, which prioritizes portal-centricity, the framework introduced in this study focuses more on users, which affects the final portal ranking.

A study [Nik21] that overlaps the most in terms of tested portals (the study does not test the Qatari, Saudi Arabian and Kuwaiti portals) shows that the portals of *France, the Netherlands, Luxemburg, Estonia, and Portugal* are top-performers in the real-time data dimension. The presented framework partially covers those and other closely related aspects in the *"Data quality"* dimension. In the evaluations conducted for this study, *France* obtained the highest possible score in that dimension, *Luxemburg and Estonia* - around the EU and total average, *the Netherlands* performed rather poorly. Variations in outcomes can be accounted for by disparities in the criteria utilized to access particular tested aspects and the ongoing portal changes.

Another study [ML17], employing the framework on top of which the framework for the present study is partially built, provided a result ranking where *Austria, France, and Croatia* are among the top performers. Although *Austria* didn't end up in the top-10 ranking of the current study, all 3 portals are high-performers. A criticism of the portals mentioned in 2017 but staying valid to this day, is that the quality of open data portals is affected by the version of the data management system [ML17]. The *Portuguese* portal inherits a lot of common features with the *French* portal by using the *udata* platform [Eta24], maintained by the French public agency Etalab. That advantage plays a clear role in the high score of the *Portuguese* portal.

A revised version of the above framework, the transparency-by-design framework [LN21], has been used in [LNL+22] to test the maturity of transparency of smart city portals. Although this framework focuses more on transparency rather than usability, missing some current trends in the resilience and sustainability aspects of OGD portals, it also propagates the idea that portals should be adapted to users spanning from those with a very limited set of skills and low digital literacy to advanced users/experts; they should facilitate data exploration and encourage the re-use of the datasets [LNL+22]. That and the current study identify concerns regarding data quality and the absence of data quality monitoring.

By comparing the findings of the present and previous research, it is evident that the current framework incorporates dimensions and sub-dimensions that effectively identify high-performing portals. However, it also offers an alternative viewpoint or sheds light on additional features that could enhance the usability and convenience of the portal. The findings of applying the framework to the OGD portals underscore the growing trend of exposing data quality metrics and advocate for enhanced communication channels between users and portal representatives. The total absence of gamification in the portals was discovered by addressing gaps identified in previous research. Emphasizing the

existing lack of gamification elements and the argument that incorporating gamification could encourage users to engage with the portal more frequently [SZCH22] may advise portal providers to incorporate such functionalities in order to attract a greater number of visitors. Aligning with previous research [NM21, ALM+18, ES23], this study confirms the need for enhanced English language support in the user interface, including the data search capabilities, as well as promoting HVD and introducing means to visualize and analyze the datasets.

6.1 Recommendations

Results of the qualitative and quantitative analyses performed during this study with subsequent identification of the best practices we came across studying selected countries, i.e., those that can be considered key to success, allow us to define them into (high-level) recommendations. Those recommendations might interest a diverse array of stakeholders, spanning government agencies, developers, researchers, and data providers. Government agencies and developers seeking to make their portals more user-friendly, collaborative, sustainable, and robust stand to benefit by implementing these recommendations to optimize their open data initiatives, fostering greater citizen engagement. Data providers who aim to enhance citizen engagement may engage in direct communication with data consumers, share vulgarized content, and maintain schema descriptions. The aforementioned recommendations highlight contemporary obstacles that open data portals face, which researchers may find valuable for their own investigations. *Nineteen* recommendations spanning *nine* sub-dimensions of the developed framework were defined.

R1: *provide full English language support ("Multilingualism").* Portals should provide translations for the navigation elements and the dataset metadata, articles, manuals, and documentation, presumably leading to greater English search support, which will facilitate the interaction of an international audience with the portal. Exemplary illustrations can be found in the portals of Bahrain, Estonia, Ireland, Malta, Oman, Qatar, Saudi Arabia, and UAE. Additionally, the European data portal provides an example of dataset metadata translation.

R2: *provide intuitive navigation ("Navigation").* Navigation elements should be easy to notice, simple and intuitive. Exemplary illustrations can be found in the portals of Poland, France, Portugal, Saudi Arabia, Czechia, Germany, and Italy.

R3: *be consistent ("Navigation", "General performance").* The portal functionalities should stay consistent, particularly regarding navigation and general performance; however, they may be applied generally. Exemplary illustrations can be found in the portals of Austria, Bahrain, Belgium, Croatia, Finland, France, Estonia, Italy, Latvia, Lithuania, Luxembourg, Netherlands, Poland, Portugal, Slovenia, Spain, and Sweden.

R4: *vulgarize content ("Data understandability").* Content vulgarization refers to transforming raw data into user-friendly formats and visualizations, making it easier for the general public to understand and utilize the information. That includes creating impact

sections and success stories, as well as displaying use cases, data reuses, applications, and services built on open data. Exemplary illustrations can be found in the portals of the Netherlands, France, Portugal, and Germany.

R5: *provide diverse means to visualize data ("Data understandability").* Diverse data visualizations encompass a range of graphical representations used to present data, including charts, graphs, maps, and interactive displays designed to convey complex information in easily understandable and engaging formats. Exemplary illustrations can be found in the portals of Qatar, Bahrain, Saudi Arabia, Lithuania, Spain, and Poland.

R6: *promote HVD ("Data understandability").* HVD promotion involves actively showcasing and making easily accessible datasets valuable for research, decision-making, or public interest to encourage their utilization and maximize their impact [NRC+23]. That might include (a) additional filtering criteria, (b) promotion on top of the catalog page, (c) a featured list, or (d) reports highlighting the most valuable datasets. Exemplary illustrations can be found in the portals of Ireland (a), Lithuania (a), Slovenia (a), Poland (a, b), the Netherlands (c), and France (d).

R7: *introduce data quality indicators ("Data quality").* Introducing data quality indicators involves establishing metrics or criteria to assess the reliability, accuracy, completeness, and consistency of data, helping users understand the trustworthiness and usability of datasets for their intended purposes. Exemplary illustrations can be found in the portals of France and Portugal.

R8: *expose data schema descriptions ("Data quality").* Exposing data schema descriptions entails providing detailed explanations of the structure and meaning of the various elements within a dataset. This facilitates comprehension and usage by those unfamiliar with the dataset's underlying structure and terminology. Exemplary illustrations can be found in the portals of Bahrain and Qatar.

R9: *expose API/SPARQL endpoints ("Data findability").* Exposing API/SPARQL endpoints facilitates seamless data access and querying, enabling programmatic retrieval and data manipulation for various applications and analyses. Exemplary illustrations can be found in the portals of Ireland, Luxembourg, France, Austria and Poland.

R10: *support complex search prompts ("Data findability").* Supporting long and complex search prompts empowers users to refine their queries with advanced filters and operators [KKI+17], facilitating precise data retrieval tailored to their needs and enhancing the overall search experience. Exemplary illustrations can be found in the portals of Ireland, Portugal, and Sweden.

R11: *implement featured topics ("Data findability").* Implementing featured topics highlights curated collections of datasets, enabling users to discover and explore relevant content in various areas quickly. Exemplary illustrations can be found in the portals of France, Italy, the Netherlands, Poland, and Luxembourg.

R12: *adopt advanced gamification ("Public engagement").* Integrating advanced gamification elements may enhance engagement, motivation, and participation among

users [SZCH22]. Competitions, quizzes, rewards, and badges are examples of advanced gamification elements.

R13: *notify about search results' updates ("Public engagement").* Creating an opportunity to subscribe to dataset or search result updates involves alerting users when new or relevant information becomes available, ensuring they stay informed and can access the latest data or content of interest. Exemplary illustrations can be found in the Polish portal.

R14: *provide comment sections ("Feedback mechanisms & service quality").* Providing comment sections allows users to engage with content by sharing their thoughts, opinions, and feedback, fostering discussions and stakeholder interaction around the data presented [ALM+18]. Exemplary illustrations can be found in the portals of France, Croatia, Lithuania, and Luxembourg.

R15: *offer diverse specialized communication channels ("Feedback mechanisms & service quality").* Offering multiple communication channels enables users to give feedback or receive help through various means such as email, issue-specific forms, chat support, and forums, accommodating diverse preferences and accessibility needs. Exemplary illustrations can be found in the portals of Saudi Arabia, Spain, and Croatia.

R16: *introduce usefulness assessment ("Feedback mechanisms & service quality").* Usefulness assessment, which can include upvoting or utilizing a 5-scale assessment, involves allowing users to provide feedback on the value of content (e.g., dataset pages, blogs, documentation pages, comments, forum posts) by expressing their ratings based on relevance, usability, and satisfaction. Exemplary illustrations can be found in the portals of Croatia, the Netherlands, Estonia, France, Portugal, and Latvia.

R17: *provide dataset release tracking ("Portal sustainability and collaboration").* Dataset release tracking involves monitoring and documenting the publication of new datasets to keep users informed about the availability of fresh data for analysis, research, or other purposes. Exemplary illustrations can be found in the French portal.

R18: *provide user satisfaction surveys ("Portal sustainability and collaboration").* Providing user satisfaction surveys entails offering structured questionnaires or feedback forms to gather valuable insights to improve user experience and address concerns or issues. Exemplary illustrations can be found in the portals of Finland, France, and Saudi Arabia.

R19: *emphasize on collaboration ("Portal sustainability and collaboration").* Initiating and advertising collaboration with other portals from the same or different country or region involves establishing partnerships and sharing resources and data to enhance the value and reach of both platforms. Merely advertising the presence of collaboration is not enough; users must also perceive how it enhances their experience of utilizing the portal. Exemplary illustrations can be found in the portals of Qatar, advertisements can be found in the portals of Germany, Austria, and Ireland.

6.2 Implications and future research

The study has the following theoretical implications: (1) it proposes an update to the criteria and metrics used to assess the user-centricity of open data portals; (2) it identifies patterns and trends observed in the literature; (1, 2) results in (3) the incorporation of those concepts and criteria into a newly developed integrated framework that compares portals from various regions and fulfills the requirement for an updated viewpoint on benchmarking open data portal performance; (4) it provides insights into the state of understudied OGD portals of the GCC states and a new perspective on the state of EU OGD portals. The practical implications are the qualitative and quantitative analyses obtained by assessing the portals and conducting the cluster analysis, as well as the defined recommendations that the portal stakeholders can use to develop user-friendly, collaborative, robust, and sustainable portals.

Future research agenda may include the expansion of the framework dimensions or sub-dimensions related to artificial intelligence (AI), which may include AI-powered assistants, recommender systems, Natural language Processing (NLP), or Large Language Model (LLM) capabilities for advanced search, dataset exploration and comprehension purposes, automation of processes, data quality improvements. Quantitatively evaluating the extent to which language barriers obstruct the interaction of users with varying degrees of digital literacy with foreign data portals could prove beneficial in proving the necessity of having full support of international languages in the portal. Lastly, conducting evaluations of portals from different regions or periodically reproducing the assessment process will enable the monitoring of new developments and trends. This may contribute to a more comprehensive understanding of the intricacies of establishing an open data ecosystem at national and interregional levels.

7 Limitations

The limitations of this research are related to (1) the framework, (2) the assessment criteria and the weighing system, (3) the sampling of datasets for assessing sub-dimensions, and (4) the assessment process.

The combinations of sub-dimensions into dimensions are practically limitless. Compiling dimensions according to pre-existing frameworks presents certain difficulties: two frameworks that serve similar purposes may possess identical sub-dimensions in distinct dimensions. In order to avoid conflating distinct approaches to incorporating sub-dimensions into dimensions, the transparency-by-design framework [LN21] was selected as the foundation for the framework created in this study. The second challenge was determining the optimal number of sub-dimensions, making the framework appealing for future applications and not overly general, thereby minimizing subjectivity in the evaluation. There were suggestions to add a check for language support in the

documentation and manuals for the "Multilingualism" dimension, check for specific data visualization types in the "Data understandability" dimension, or a separate portal suggestions form in the "Feedback mechanisms and service quality" dimension. They were excluded because they are components of other sub-dimensions, and the intention was to avoid having such granular evaluation criteria.

With the intention of keeping the framework appealing for future applications, the boolean assessment is predominantly used. That required building the criteria so that it could be answered with yes/no. That limits the expression of a third option, which was decided to be expressed as an additional note, resulting in a qualitative analysis. The selection of the weighting system was determined by the need to avoid an overly simplistic or complex system.

In relation to the sampling of the datasets, it must be acknowledged that the sampling is exploratory in nature and can be seen as insufficiently representative. The number of fourteen datasets included in the sample was determined to be optimal since it ensures an equitable opportunity for data exploration while also meeting the constraints of time and human resources.

Time and human resource limitations restricted the evaluation of each portal to a single instance. Furthermore, due to the variations in feature implementation across the various portals, the automated evaluation of the portals was restricted to the scenarios outlined in the Methodology section.

8 Conclusion

This study develops an integrated usability framework for evaluating open data portals that focuses on: (1) the portal's ability to adapt to a diverse user base; (2) promoting user collaboration and participation; and (3) enabling users to understand and explore the data. The framework, consisting of 72 dimensions, was developed and then applied to 33 EU and GCC OGD portals to assess their usability. As a result of its application, (a) each portal was assessed, (b) statistics about the portal performances were gathered, (c) portals were ranked, (d) best practices and pain points for portals were derived, (e) trends in portal design and collaborative initiatives between portals were identified, (f) cluster analysis based on the score matrix was conducted. While assessing each portal, the sub-dimension, dimension, and total score were computed using the weighing system. The portal performance statistics included the average scores for the EU and GCC portals and the identification of top and low-performers. Portals were ranked according to the results gained in individual dimensions, as well as by the total score. In conjunction with the quantitative analysis, best practices and pain points were identified in the qualitative analysis using portal-specific examples. Two types of cluster analyses (K-means and hierarchical) were conducted in order to investigate the similarities among different portals and determine which clusters provide noteworthy performance examples across various dimensions. Trends in portal design and collaborative initiatives between portals at the intra- and interregional levels were identified which were then transformed into 19 recommendations.

The high performance of top European portals (based on EU Open Data Maturity Reports) within this framework can be seen as an indicator of some degree of consistency between the proposed framework and existing widely used benchmarks and indices. However, an alternative view of these portals has been presented based on the unique aspects/dimensions that the presented framework considers. Additionally, it was discovered that the portals of Saudi Arabia, Qatar, and Bahrain are competent and even trend-setting in certain sub-dimensions.

The presented analysis suggests that portals should focus on improving the multilingual experience, allowing users to communicate their needs to portal representatives or data publishers. It should also make it easier for users to understand how to use datasets and find them on the portal.

It has been observed that there is a growing trend of exposing data quality indicators. Additionally, involving users in the portal ecosystem results in a more vibrant and engaging experience, increasing the likelihood of repeated use of the portal even for newcomers, making them part of the open data ecosystem. A common problem is that portals often fail to adequately highlight implemented features, which are often hidden and difficult to find, not to say mention the lack of features that a user might expect, including assistants, AI-augmented recommender systems, advanced search or NLP or LLM capabilities for advanced search or examining datasets, as well as gamification

elements. Implementing effective feedback mechanisms can enhance public participation and elevate the general quality of the datasets or portals. This framework should be revisited once there are examples of how the above technologies can be advantageous for open data portals. Currently, the framework can be seen as the "minimum set of requirements" that the OGD portal must comply with.

The analysis results indicate that despite the lack of research comparing portals of different regions, including the EU and GCC, conducting a comparative analysis of portals from different regions is feasible and advisable. While portals from the same region often have similar strengths and weaknesses, studying portals from different regions can provide new insights into how the same features may be implemented differently.

Although cross-border and inter-regional portal collaboration is not widely observed, several examples have been found in both regions. The Qatar portal represents an innovative example of cross-regional innovation, while the German and Austrian portals are promoting their cooperation with the Swiss and Lichtenstein portals, thereby moving towards an enhanced and more sustainable open data ecosystem (among Germany, Austria, Switzerland, and Liechtenstein). While the importance of collaborations in creating a more connected open data ecosystem is often underestimated, these efforts are critical to progressing toward this goal.

The results of this study have been accepted for presentation at the 25th Annual International Conference on Digital Government Research (DGO 2024) and will be published in the Association for Computing Machinery Digital Library (ACM DL) [MN24] with an extended version submitted to a journal.

References

[ABAK23] Riyadh Al-Balushi and Al-Waleed Al-Kiyumi. Open data oman (community version), 2023.

[Acc23] AccessibilityChecker.org. Accessibility checker - eaa compliance, 2023.

[ADAD17] Eric Afful-Dadzie and Anthony Afful-Dadzie. Open Government Data in Africa: A preference elicitation analysis of media practitioners. *Government Information Quarterly*, 34(2):244–255, April 2017.

[ALM+18] Charalampos Alexopoulos, Euripidis Loukis, Spiros Mouzakitis, Michalis Petychakis, and Yannis Charalabidis. Analysing the Characteristics of Open Government Data Sources in Greece. *Journal of the Knowledge Economy*, 9(3):721–753, September 2018.

[AODucDPh22] A. Abella, M. Ortiz-De-urbina criado, and C. De-Pablos-heredero. Criteria for the identification of ineffective open data portals: pretender open data portals. *Profesional de la Informacion*, 31(1), 2022.

[AZS21] Ali Pirannejad Anneke Zuiderwijk and Iryna Susha. Comparing open data benchmarks: Which metrics and methodologies determine countries' positions in the ranking lists? *Telematics and Informatics*, 62:101634, 2021.

[BBV22] Raissa Barcellos, Flavia Bernardini, and José Viterbo. Towards defining data interpretability in open data portals: Challenges and research opportunities. *Information Systems*, 106:101961, May 2022.

[Car22] Carsaniga, Giulia and Lincklaen Arriëns, Eline N. and Dogger, Jochem and van Assen, Mariska and Cecconi, Gianfranco. *Open data maturity report 2022*. Publications Office of the European Union., LU, 2022.

[Com] European Commision. European Commission to make public sector high-value datasets openly available | data.europa.eu.

[Com23] European Commision. Open data | Shaping Europe's digital future, February 2023.

[CSO+20] U.C. Chua, K.L. Santiago, I.B.M. Ona, R.M.N. Peña, G.Z.S. Marasigan, P.G.A.D. Reyes, and B.P.V. Samson. From Access to Effective Use: Open Data Portals for Everyday Citizens. *AsianCHI '20: Proceedings of the 2020 Symposium on Emerging Research from Asia and on Asian Contexts and Cultures*, pages 61–64, 2020.

[Cut20] Martin Cutts. *Oxford Guide to Plain English.* February 2020.

[DLC18] Kawtar Younsi Dahbi, Hind Lamharhar, and Dalila Chiadmi. Exploring dimensions influencing the usage of Open Government Data portals. In *Proceedings of the 12th International Conference on Intelligent Systems: Theories and Applications*, SITA'18, pages 1–6, New York, NY, USA, October 2018. Association for Computing Machinery.

[ES23] E. Elsawy and A. Shehata. Open government data initiatives in the Maghreb countries: An empirical analysis. *IFLA Journal*, 49(1):61–73, 2023.

[Eta24] Etalab. Udata platform, 2024.

[FDK20] M. Ferati, F. Dalipi, and Z. Kastrati. Open government data through the lens of universal design. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12189 LNCS:331–340, 2020. ISBN: 9783030491079.

[GC17] Mark Gill and Jon Corbett. Downscaling: Understanding the influence of open data initiatives in smaller and mid-sized cities in British Columbia, Canada. *Canadian Geographies / Géographies canadiennes*, 61(3):346–359, 2017. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cag.12372.

[GPC23] Rikke Magnussen Georgios Papageorgiou, Loukis Euripides and Yannis Charalabidis. Open data journalism: a domain mapping review. In *In Proceedings of the 16th International Conference on Theory and Practice of Electronic Governance, ACM International Conference Proceeding Series*, pages 159–166, 2023.

[GS20] P. Govender and V. Sivakumar. Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmospheric Pollution Research*, 11(1):40–56, January 2020.

[Hev07] Alan R Hevner. A Three Cycle View of Design Science Research. 19, 2007.

[HLR+21] Peter W Harrison, Rodrigo Lopez, Nadim Rahman, Stefan Gutnick Allen, Raheela Aslam, Nicola Buso, Carla Cummins, Yasmin Fathy, Eloy Felix, Mihai Glont, Suran Jayathilaka, Sandeep Kadam, Manish

Kumar, Katharina B Lauer, Geetika Malhotra, Abayomi Mosaku, Ossama Edbali, Young Mi Park, Andrew Parton, Matt Pearce, Jose Francisco Estrada Pena, Joseph Rossetto, Craig Russell, Sandeep Selvakumar, Xènia Pérez Sitjà, Alexey Sokolov, Ross Thorne, Marianna Ventouratou, Peter Walter, Galabina Yordanova, Amonida Zadissa, Guy Cochrane, Niklas Blomberg, and Rolf Apweiler. The COVID-19 Data Portal: accelerating SARS-CoV-2 and COVID-19 research through rapid open access data sharing. *Nucleic Acids Research*, 49(W1):W619–W623, July 2021.

[HMGS20] Camila Andrea Herrera-Melo and Juan Sebastián González-Sanabria. Proposal for the Evaluation of Open Data Portals. *Revista Facultad de Ingeniería*, 29(54):e10194–e10194, 2020. Number: 54.

[HR23] Augusto Herrmann and Tod Robbins. Dataportals.org, 2023.

[JCZ12] M. Janssen, Y. Charalabidis, and A. Zuiderwijk. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29(4):258–268, 2012.

[KAKD22] K. Kević, C. Alexopoulos, and A. KuveÅ3/4dić Divjak. Assessing Information and System Quality of Open Government Data Portals in Croatia. *ICEGOV '22: Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance*, pages 316–324, 2022.

[Kao23] Hao-En Kao. Between international practice and academia: Review and integration of open government data benchmarks. In *In Proceedings of the 24th Annual International Conference on Digital Government Research*, pages 73–89, 2023.

[KH21] Dewi Krismawati and Achmad Nizar Hidayanto. The User Engagement of Open Data Portal. In *2021 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pages 1–6, October 2021. ISSN: 2330-4588.

[Kit04] Barbara Kitchenham. Procedures for performing systematic reviews. Technical report, Keele University, 2004.

[KKI+17] Emilia Kacprzak, Laura M. Koesten, Luis-Daniel Ibáñez, Elena Simperl, and Jeni Tennison. A Query Log Analysis of Dataset Search. In Jordi Cabot, Roberto De Virgilio, and Riccardo Torlone, editors, *Web Engineering*, Lecture Notes in Computer Science, pages 429–436, Cham, 2017. Springer International Publishing.

[KKI+19] Emilia Kacprzak, Laura Koesten, Luis-Daniel Ibáñez, Tom Blount, Jeni Tennison, and Elena Simperl. Characterising dataset search—An analysis of search logs and data requests. *Journal of Web Semantics*, 55:37–55, March 2019.

[Kli19] Jakub Klimek. DCAT-AP representation of Czech National Open Data Catalog and its impact. *Journal of Web Semantics*, 55:69–85, March 2019.

[KM24] K-Monitor. Open data hungary (community version), 2024.

[Kni20] Laura van Knippenberg. *Open data maturity report 2020*. Publications Office of the European Union, Luxembourg, 2020. OCLC: 1242741263.

[KRN+18] S. Kubler, J. Robert, S. Neumaier, J. Umbrich, and Y. Le Traon. Comparison of metadata quality in open data portals using the Analytic Hierarchy Process. *Government Information Quarterly*, 35(1):13–29, 2018.

[LMV+21] M. Lněnička, R. Machova, J. Volejníková, V. Linhartová, R. Knezackova, and M. Hub. Enhancing transparency through open government data: the case of data portals and their features and capabilities. *Online Information Review*, 45(6):1021–1038, 2021.

[LN21] Martin Lnenicka and Anastasija Nikiforova. Transparency-by-design: What is the role of open data portals? *Telematics and Informatics*, 61:101605, August 2021.

[LNL+22] Martin Lnenicka, Anastasija Nikiforova, Mariusz Luterek, Otmane Azeroual, Dandison Ukpabi, Visvaldis Valtenbergs, and Renata Machova. Transparency of open data ecosystems in smart cities: Definition and assessment of the maturity of transparency in 22 smart cities. *Sustainable Cities and Society*, 82:103906, July 2022.

[LNL+24a] Martin Lnenicka, Anastasija Nikiforova, Mariusz Luterek, Petar Milic, Daniel Rudmark, Sebastian Neumaier, Caterina Santoro, Cesar Casiano Flores, Marijn Janssen, and Manuel Pedro Rodríguez Bolívar. Identifying patterns and recommendations of and for sustainable open data initiatives: A benchmarking-driven analysis of open government data initiatives among European countries. *Government Information Quarterly*, 41(1):101898, March 2024.

[LNL+24b] Martin Lnenicka, Anastasija Nikiforova, Mariusz Luterek, Petar Milic, Daniel Rudmark, Sebastian Neumaier, Caterina Santoro, Cesar Casiano Flores, Marijn Janssen, and Manuel Pedro Rodríguez Bolívar. Identifying patterns and recommendations of and for sustainable open data initiatives: a benchmarking-driven analysis of open government data initiatives among european countries. *Government Information Quarterly*, 41:101898, 2024.

[Mad] Tambiama Madiega. The european data act.

[MAL+21] Ibrahim Mutambik, Abdullah Almuqrin, John Lee, Justin Zuopeng Zhang, Abdulaziz Alomran, Taha Omar, Ahmad Floos, and Abdullah Homadi. Usability of the G7 Open Government Data Portals and Lessons Learned. *Sustainability*, 13(24):13740, December 2021.

[Mei19] K.L. Meissner. Leveraging interregionalism: EU strategic interests in Asia, Latin America and the Gulf region. *International Politics*, 56(3):359–374, 2019.

[MHL18] Renáta Máchová, Miloslav Hub, and Martin Lnenicka. Usability evaluation of open data portals: Evaluating data discoverability, accessibility, and reusability from a stakeholders' perspective. *Aslib Journal of Information Management*, 70(3):252–268, January 2018. Publisher: Emerald Publishing Limited.

[MJJ21] R. Matheus, M. Janssen, and T. Janowski. Design principles for creating digital transparency in government. *Government Information Quarterly*, 38(1), 2021.

[ML17] Renata Máchová and Martin Lnénicka. Evaluating the Quality of Open Data Portals on the National Level. *Journal of theoretical and applied electronic commerce research*, 12(1):21–41, 2017. Publisher: Universidad de Talca.

[MLN22] Mariusz Luterek Martin Lnenicka and Anastasija Nikiforova. Benchmarking open data efforts through indices and rankings: Assessing development and contexts of use. *Telematics and Informatics*, 66:101745, 2022.

[MN24] Fillip Molodtsov and Anastasija Nikiforova. An Integrated Usability Framework for Evaluating Open Government Data Portals: Comparative Analysis of EU and GCC Countries. In *Proceedings of the 25th Annual International Conference on Digital Government Research*. Association for Computing Machinery, 2024.

[MTKK18] Keegan McBride, Maarja Toots, Tarmo Kalvet, and Robert Krimmer. Open Government Data Driven Co-creation: Moving Towards Citizen-Government Collaboration. *International Conference on Electronic Government*, 11020:184–195, 2018. Series Title: Lecture Notes in Computer Science.

[Nik20a] A. Nikiforova. Assessment of the usability of latvia's open data portal or how close are we to gaining benefits from open data. *Journal of Scientific and Industrial Research*, pages 51–58, 2020.

[Nik20b] Anastasija Nikiforova. Comparative analysis of national open data portals or whether your portal is ready to bring benefits from open data. pages 81–88, 2020.

[Nik21] Anastasija Nikiforova. Smarter Open Government Data for Society 5.0: Are Your Open Data Smart Enough? *Sensors*, 21(15):5204, January 2021. Number: 15 Publisher: Multidisciplinary Digital Publishing Institute.

[NL21] Anastasija Nikiforova and Martin Lnenicka. A multi-perspective knowledge-driven approach for analysis of the demand side of the Open Government Data portal. *Government Information Quarterly*, 38(4), 2021.

[NM90] Jakob Nielsen and Rolf Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90, pages 249–256, New York, NY, USA, March 1990. Association for Computing Machinery.

[NM21] Anastasija Nikiforova and Keegan McBride. Open government data portal usability: A user-centred usability analysis of 41 open government data portals. *Telematics and Informatics*, 58, 2021.

[NRC+23] Anastasija Nikiforova, Nina Rizun, Magdalena Ciesielska, Charalampos Alexopoulos, and Andrea Miletič. Towards High-Value Datasets determination for data-driven development: a systematic literature review, 2023.

[NTM19] Fatima W. Nosheen, Rasool Tayyaba, and Sajid Muhammad. Challenges to Use Open Government Data through the Citizens' Lens: A Systematic Review. *Journal of Political Studies*, July 2019.

[Oba09] Barack Obama. Open Government Directive, 2009.

[Pag23] Page, Martin and Hajduk, Emir and Lincklaen Arriëns, Eline N. and Cecconi, Gianfranco and Briknhuis, Suzan. *Open data maturity report 2023*. Publications Office of the European Union., LU, 2023.

[PZJ20a] A. Purwanto, A. Zuiderwijk, and M. Janssen. Citizen engagement with open government data: Lessons learned from Indonesia's presidential election. *Transforming Government: People, Process and Policy*, 14(1):1–30, 2020.

[PZJ20b] A. Purwanto, A. Zuiderwijk, and M. Janssen. Citizens' trust in open government data. In *ACM International Conference Proceeding Series*, pages 310–318, 2020.

[PZJ20c] Arie Purwanto, Anneke Zuiderwijk, and Marijn Janssen. Citizen Engagement with Open Government Data: A Systematic Literature Review of Drivers and Inhibitors. *International Journal of Electronic Government Research*, 16(3):1–25, 2020.

[QDM19] Alfonso Quarati and Monica De Martino. Open government data usage: a brief overview. In *Proceedings of the 23rd International Database Applications & Engineering Symposium*, IDEAS '19, pages 1–8, New York, NY, USA, June 2019. Association for Computing Machinery.

[Reg20] Luigi Reggi. Open government data strategies and governance networks inclusiveness: Lessons about transparency, use, and the creation of ecosystems. *State University of New York at Albany, 2020*, pages 1–149, 2020.

[RVCKP21] Vigan Raca, Goran Velinov, Betim Cico, and Margita Kon-Popovska. Application-based Framework for Analysis, Monitoring and Evaluation of National Open Data Portals. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 12(11), January 2021. Number: 11 Publisher: The Science and Information (SAI) Organization Limited.

[Sax18a] S. Saxena. Open government data (OGD) in six Middle East countries: an evaluation of the national open data portals. *Digital Policy, Regulation and Governance*, 20(4):310–322, 2018.

[Sax18b] Stuti Saxena. National open data frames across Japan, The Netherlands and Saudi Arabia: role of culture. *foresight*, 20(1):123–134, January 2018. Publisher: Emerald Publishing Limited.

[Sax19] Stuti Saxena. Evaluation of the National Open Government Data (OGD) Portal of Saudi Arabia. In Anna Visvizi and Miltiadis D. Lytras, editors, *Politics and Technology in the Post-Truth Era*, Emerald Studies in Politics and Technology, pages 221–235. Emerald Publishing Limited, January 2019.

[SGJ15] Iryna Susha, Åke Grönlund, and Marijn Janssen. Organizational measures to stimulate user engagement with open data. *Transforming Government: People, Process and Policy*, 9(2):181–206, January 2015.

[SGLPM18] Raffaele Sisto, Javier García López, José Manuel Paéz, and Elena Maté Múgica. Open Data Assessment in Italian and Spanish Cities. In Adriano Bisello, Daniele Vettorato, Pierre Laconte, and Simona Costa, editors, *Smart and Sustainable Planning for Cities and Regions*, Green Energy and Technology, pages 121–131, Cham, 2018. Springer International Publishing.

[SJ15] Renee E. Sieber and Peter A. Johnson. Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly*, 32(3):308–315, July 2015.

[SM18] Thomas Schauppenlehner and Andreas Muhar. Theoretical Availability versus Practical Accessibility: The Critical Role of Metadata Management in Open Data Portals. *Sustainability*, 10(2):545, February 2018. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[SMKV23] S. Sheoran, S. Mohanasundaram, R. Kasilingam, and S. Vij. Usability and Accessibility of Open Government Data Portals of Countries Worldwide: An Application of TOPSIS and Entropy Weight Method. *International Journal of Electronic Government Research*, 19(1), 2023.

[SVM22] Žaklina Spalević, Nataša Veljković, and Petar Milić. QODA – Methodology and Legislative Background for Assessment of Open Government Datasets Quality. *Journal of Scientific & Industrial Research*, 82(1):38–49, December 2022. Number: 1.

[SZC+15] I. Susha, A. Zuiderwijk, Y. Charalabidis, P. Parycek, and M. Janssen. Critical factors for open data publication and use: A comparison of city-level, regional, and transnational cases. *eJournal of eDemocracy and Open Government*, 7(2):94–115, 2015.

[SZCH22] A. Simonofski, A. Zuiderwijk, A. Clarinval, and W. Hammedi. Tailoring open government data portals for lay citizens: A gamification

theory approach. *International Journal of Information Management*, 65, 2022.

[SZJG15] Iryna Susha, Anneke Zuiderwijk, Marijn Janssen, and Ake Grönlund. Benchmarks for Evaluating the Progress of Open Data Adoption: Usage, Limitations, and Lessons Learned. *Social Science Computer Review*, 33(5):613–630, 2015.

[Ver24] Stefaan G. Verhulst. Are we entering a "data winter"?, February 2024.

[WCR18] D. Wang, C. Chen, and D. Richards. A prioritization-based analysis of local open government data portals: A case study of Chinese province-level governments. *Government Information Quarterly*, 35(4):644–656, 2018.

[WIK+17] V. Weerakkody, Z. Irani, K. Kapoor, U. Sivarajah, and Y.K. Dwivedi. Open data and its usability: an empirical view from the Citizen's perspective. *Information Systems Frontiers*, 19(2):285–300, 2017.

[WRC18] Di Wang, Deborah Richards, and Chuanfu Chen. An Analysis of Interaction Between Users and Open Government Data Portals in Data Acquisition Process. In Kenichi Yoshida and Maria Lee, editors, *Knowledge Management and Acquisition for Intelligent Systems*, Lecture Notes in Computer Science, pages 184–200, Cham, 2018. Springer International Publishing.

[WS20] Victoria Wang and David Shepherd. Exploring the extent of openness of open government data – A critique of open government datasets in the UK. *Government Information Quarterly*, 37(1):101405, January 2020.

[ZCS21] Anneke Zuiderwijk, Yu-Che Chen, and Fadi Salem. Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly*, 38(3):101577, July 2021.

[ZF19] Xiaohua Zhu and Mark Antony Freeman. An evaluation of U.S. municipal open data portals: A user interaction framework. *Journal of the Association for Information Science and Technology*, 70(1):27–37, 2019.

[ZJD14] A. Zuiderwijk, M. Janssen, and C. Davis. Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, 19(1-2):17–33, 2014.

[ZJD15] Anneke Zuiderwijk, Marijn Janssen, and Yogesh K. Dwivedi. Acceptance and use predictors of open data technologies: Drawing upon the unified theory of acceptance and use of technology. *Government Information Quarterly*, 32(4):429–440, October 2015.

[ZJSS22] Wenli Zhang, Hongbo Jiang, Qigan Shao, and Ting Shao. Construction of the Evaluation Model of Open Government Data Platform: From the Perspective of Citizens' Sustainable Use. *Sustainability*, 14(3):1415, January 2022. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

[ZPS21] A. Zuiderwijk, A. Pirannejad, and I. Susha. Comparing open data benchmarks: Which metrics and methodologies determine countries' positions in the ranking lists? *Telematics and Informatics*, 62, 2021.

[ZTMA21] X. Zhu, C. Thomas, J.C. Moore, and S. Allen. Open Government Data Licensing: An Analysis of the U.S. State Open Government Data Portals. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12646 LNCS:260–273, 2021. ISBN: 9783030713041.

Appendix

I. Clustering code listing

```
import pandas as pd
from sklearn.cluster import KMeans
from scipy.cluster.hierarchy import dendrogram, linkage, fcluster
import matplotlib.pyplot as plt

scores_file = 'file.csv' # Replace with your actual file path

data = pd.read_csv(scores_file)

# Selecting the relevant columns for clustering (excluding the
    country column)
X = data.iloc[:, 1:]

# Determining the optimal number of clusters using the Elbow method
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300,
        n_init=10, random_state=0)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)

# Plotting the Elbow Method graph
plt.figure(figsize=(10, 5))
plt.plot(range(1, 11), wcss)
plt.title('Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')  # Within cluster sum of squares
plt.savefig('elbow.png')
plt.show()

# Choose the number of clusters based on the Elbow graph
n_clusters = 4

# Performing k-Means clustering

kmeans = KMeans(n_clusters=n_clusters, init='k-means++', max_iter
    =300, n_init=10, random_state=0)

# Fitting the model to the data and predicting clusters
clusters_k = kmeans.fit_predict(X)

# Adding the cluster information to the original dataset
data['Cluster_' + str(n_clusters)] = clusters_k
```

```
# Displaying the countries with their cluster labels
clustering_table = data[['country', 'Cluster_' + str(n_clusters)]]
clustering_table.to_csv('kmeans-clustering.csv',index=False)
clustering_table

# Hierarchical clustering
# Load the dataset
data = pd.read_csv(scores_file)

# Selecting the relevant columns for clustering (excluding the
    country column)
X = data.iloc[:, 1:]

# Generating the linkage matrix for hierarchical clustering using the
    Ward method
Z = linkage(X, method='ward')

# Plotting the dendrogram
plt.figure(figsize=(18, 18))
plt.title('Hierarchical Clustering Dendrogram')
plt.xlabel('Country')
plt.ylabel('Distance')
dendrogram(
    Z,
    leaf_rotation=90.,  # rotates the x axis labels
    leaf_font_size=10.,  # font size for the x axis labels
    labels=data['country'].values  # Using country names as labels
)
plt.savefig('dendrogram.png')
plt.show()

# Define a specific distance threshold
distance_threshold = 17

# Creating clusters based on the specified distance threshold
clusters = fcluster(Z, distance_threshold, criterion='distance')

# Adding the cluster information to the original dataset
data['Hierarchical_Cluster'] = clusters

# Displaying the countries with their hierarchical cluster labels
hierar_clustering_table = data[['country', 'Hierarchical_Cluster']]
hierar_clustering_table.to_csv('hierarchical-clustering.csv',index=
    False)
hierar_clustering_table
```

II. Licence

Non-exclusive licence to reproduce thesis and make thesis public

I, **Fillip Molodtsov**,
(author's name)

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to

   reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

   An Integrated Usability Framework for Evaluating Open Government Data Portals: Comparative Analysis of EU and GCC Countries,
   (title of thesis)

   supervised by Anastasija Nikiforova.
   (supervisor's name)

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Fillip Molodtsov
08/05/2024