Alexopoulos, C., Ali, M., Maratsi, M.I. *et al.* Assessing the availability and interoperability of open government data (OGD) supporting sustainable development goals (SDGs) and value creation in the gulf cooperation council (GCC). *Qual Quant* (2024). https://doi.org/10.1007/s11135-024-02025-2

Assessing the Availability and Interoperability of Open Government Data (OGD) supporting Sustainable Development Goals (SDGs) and Value Creation in the Gulf Cooperation Council (GCC)

Charalampos Alexopoulos, University of the Aegean(alexop@aegean.gr)
Mohsan Ali, University of the Aegean(mohsan@aegean.gr)
Maria Ioanna Maratsi, University of the Aegean(ioanna.m@aegean.gr)
Nina Rizun, Gdańsk University of Technology (ninarizun@pg.edu.pl)
Yannis Charalabidis, University of the Aegean (yannisx@aegean.gr)
Euripides Loukis, University of the Aegean (eloukis@aegean.gr)
Stuti Saxena, Graphic Era University (stutisaxenaogd.vishnu@gmail.com)

Abstract

Value creation and innovation by a range of stakeholders, including citizens, analysts, journalists, non-profit entities, etc. are the hallmarks of Open Government Data (OGD) initiatives. At the same time, the availability and interoperability of datasets are determined as two of the most important factors for value creation. In parallel, the United Nations' Sustainable Development Goals (SDGs) are meant to be realized to attain quality of life through the development of initiatives based on the value creation and innovation provided by the afore-identified stakeholders, thus, the information provided from the public sector (OGD) regarding SDGs would help the monitoring of current and the identification of next actions and initiatives. The examination of SDG datasets availability and the assessment of their interoperability would provide valuable insights regarding the extent to which the data can contribute to value creation. In that vein, the availability and the interoperability dimensions of the OGD provisioned via the national OGD portals of the six Gulf Cooperation Council (GCC) constituents; viz., Bahrain, Qatar, Oman, Kuwait, United Arab Emirates (UAE) and Saudi Arabia, are studied based on a semi-automatic methodological approach. In this light, the present study seeks to investigate to what extent the national OGD portals of the GCC region cater for the interoperability dimensions, more specifically, the semantic interoperability, to facilitate value creation and innovation. Semantic interoperability dimensions were investigated via the cosine similarity calculations in Python to understand the extent to which the availability of OGD via the national OGD portals facilitates their interoperability. Findings show that the value creation and innovation initiatives to realize the SDGs' attainment is dependent upon both the availability and the extent of interoperability for all SDGs. GCC countries present different levels of both factors, which is suggestive of the mismatch of the OGD provision and their attributes resulting in low interoperability. Findings from the study are indicators that the GCC countries should develop different strategies regarding the availability and the interoperability of SDG-related OGD in order to stimulate innovation and value creation.

Keywords:

Open Government Data, OGD, Semantic interoperability, Data availability, Value creation, Sustainable Development Goals, SDGs, Gulf Cooperation Council, GCC

1. Introduction

The United Nations set forth the Sustainable Development Goals (SDGs) to improve the quality of life. These SDGs are pertinent to different aspects of life and have objectives to meet, such as eradicating poverty (SDG 1), attaining zero hunger (SDG 2), achieving health and well-being (SDG 3), improving quality of education (SDG 4), and achieving gender quality (SDG 5) are among others (UN, 2015). In the United Nations' perspective, the Sustainable Development Goals (SDGs) have the potential to contribute to the enhancement of quality of life and serve as one of their primary and global objectives until 2023. The seventeen Sustainable Development Goals (SDGs) established by the United Nations aim to enhance quality of life through various socio-economic indicators, including education, business, trade, energy, and infrastructure. Achieving these goals requires coordinated efforts across sectors and active involvement from multiple stakeholders. Undoubtedly, governments around the globe are bringing every service closer to the citizens in order to concretize the UN vision of achieving quality of life via disruptive technologies and information communication technologies (ICT) through multiple interfaces with the citizens (Ahlgren, Hidell & Ngai, 2016; Janowski, 2016; Medaglia, Misuraca & Aquaro, 2021).

One such interface pertains to that of the Open Government Data (OGD) initiatives which are undertaken to realize the overarching objectives of transparency, collaboration and participation (Alexopoulos, Spiliotopoulou & Charalabidis, 2013; Bertot et al., 2014; De Blasio & Selva, 2019; Janssen, 2011) apart from improvising the public service delivery formats (Gonzalvez-Gallego, Nieto-Torrejon & Perez-Carceles, 2020) through value creation and innovation (Jetzek, Avital & Bjorn-Andersen, 2013; Zuiderwijk, Janssen & Davis, 2014) by a range of stakeholders; journalists, software developers, professionals, academia, citizens and the like (Cho, 2023; Vetro et al., 2016). The question is that how open government data is relevant to the SDGs achievement. In the World Bank report, Open Data has the potential to be a propelling force in the pursuit of the SDGs by providing elucidating information about naturally available resources, government operations, public services, and demographic data. In addition to facilitating the identification of effective pathways to national priorities, these information resources also address national issues (Gurin, Manley & Ariss, 2015). Data is essential for driving economic growth, enhancing public services, increasing transparency and accountability, and improving government efficiency and disaster resilience. It also supports assessing the SDG goals by providing standards, tools for accountability, and evidence-oriented impact assessment (Gurin, Manley & Ariss, 2015).

Several research articles have examined the relationship between Open Government initiatives and the SDGs. Petrov et al. (2016) highlight the crucial role of open data in achieving the UN SDGs, detailing how these initiatives aim to convert publicly available information into reusable formats, potentially generating trillions in economic value annually (Petrov, Gurin & Manley, 2016). To enhance the utility of open data for the SDGs, Palacios and Piedra (2019) suggest creating semantic knowledge graphs to link dispersed datasets with specific goals (Palacios & Piedra, 2019). The Indian national open data portal provides centralized access to

SDG-related data; however, it faces critical challenges such as lack of awareness, data quality issues, and limited impact on citizens' lives (Jaiswal, 2019).

OGD refers to the provision of structural and functional administrative datasets via dedicated Web portals to be re-used by the stakeholders for value creation and innovation. OGD is license-free and the same is available across a myriad of socio-economic indices, as listed above. To ensure that value creation and innovation happen, it is important that there should be interoperability of the datasets, as well, for their optimal usage (Charalabidis et al., 2018; Janssen, Estevez & Janowski, 2014; Jimenez, Solanas & Falcone, 2014; Mareti, Russo & del Gobbo, 2021; Morando, 2013) and the extent of the datasets' interoperability is suggestive of the possibilities of cross-domain integration based on the OGD usability (Cho, 2023). In this light, further research has been warranted to assess the technical, semantic, and organizational dimensions of the interoperability of OGD (Charalabidis, Alexopoulos & Loukis, 2016).

With this landscape, the present study seeks to address the research question: "To what extent are the national OGD portals of the GCC region catering for the interoperability dimensions - specifically, the semantic interoperability- to facilitate value creation and innovation?" The rationale for picking up the GCC region i.e., Oman, Bahrain, Kuwait, Qatar, Saudi Arabia and the United Arab Emirates (UAE)- is two-fold; first of all, the authors are familiar with the region adequately and, secondly, academic interest in the GCC region has been scant as far as the OGD lens is concerned (Saxena, 2017). Given the economic significance of the region, it is important that the OGD initiatives of this region be scanned for further research.

Putting things in perspective, the present research seeks to ascertain the extent to which the necessary information regarding SDGs is provided, and, as a second step, to examine the extent to which the semantic interoperability of the datasets is facilitated via the portals. Specifically, the authors will evaluate the coverage of requirements towards the support of SDGs and then will evaluate the coverage of semantic data interoperability requirements which are reflective of the ease with which value creation and innovation may happen. Challenges in terms of interoperability of datasets are identified in line with the inferences drawn from the empirical protocols. Academic and practitioner insights constitute the concluding sections of the paper.

The objective of this study is twofold: first, it aims to unify the GCC nations (and potentially other countries) in their provision of open datasets related to the SDGs; second, it seeks to foster collaboration among nations working towards common SDGs. Additionally, this study serves as a wake-up call for countries that are not yet providing datasets related to these SDGs. By aligning open datasets with SDGs, the overarching goal of enhancing the quality of life can be more effectively achieved.

The remainder of the paper is structured as follows: Section 2 presents the background and related research in terms of data interoperability and OGD in the GCC. Section 3 provides the research methodology followed, and the study results are provided in Section 4. Section 5 includes a Discussion of the presented study, while Section 6 constitutes its concluding remarks. Finally, future research directions and practitioner implications are mentioned in the last Sections of this study.

2. Related Research

The present study reviews three strands of literature: a) literature on the interoperability of datasets, b) literature on Open Government Data (OGD) focused on the GCC region, and c) literature on the technical challenges related to OGD interoperability. These research threads, discussed in Sections 2.1, 2.2, and 2.3, are crucial for understanding the current needs and the status of OGD dataset interoperability provided through GCC portals. Based on this review, a research gap is identified in Section 2.4.

2.1. Literature referring to the interoperability of datasets

Data interoperability has been conceptualized into four types: technical (usage of HTTP on the Web helps in making the datasets technically interoperable with all the other datasets published on the web), syntactic (this is not of much consequence for the OGD users given that the OGD tapped by the users is non-proprietary and machine-processable), organizational (cultural and organizational processes that are high-level) and semantic (datasets having a common terminology wherein the same term has the same meaning or is used in the same way in the datasets) (Rezaei, Chiew & Lee, 2014). Interoperability becomes significant in the digital contexts, smart cities, for instance, because a range of stakeholders are involved in the harnessing and utilization of multiple platforms that are interspersed between data dashboards or visualization tools and to affect the integration of these heterogeneous platforms, interoperability becomes important (Chaturvedi & Kolbe, 2019; Lodato, French & Clark, 2021).

Interoperability helps in the integration of raw OGD, data sharing of services and the homogenization of data samples. Any interoperability framework is made robust by the intelligent services layer which facilitates the interaction with the citizens. Thus, interoperability results in conceiving the cross-integrated value addition services and products wherein stakeholders infer the extent to which the innovative solutions are possible across domains. Likewise, there is a multitudinous set of platforms that run across diverse domains which are fed by diverse sources. For instance, the case of mySMARTLIfe, which facilitates interoperability with the Open Specifications Framework, thereby catering to the interoperability needs of the three smart cities across different countries, France, Germany and Finland such that open APIs (Application Programming Interfaces), OGD and open SDKs (Software Development Kits) are being factored into account (Hernandez et al., 2020).

The last-mentioned research underlines interoperability as foundational (facilitates data exchange between two information technologies), structural (providing the defining characteristics in terms of the structure of format of data exchange with emphasis upon the syntax of data exchange) and semantic (many systems are able to comprehend the data because of the common data models). Furthermore, interoperability plays a key role on the Internet of Things (IoT) technologies like cloud platforms, sensors, high-speed networks and the like (Ahlgren, Hidell & Ngai, 2016). Likewise, interoperability in the context of the OGD assumes significance for the combinations and analyses resulting from the multitudinous datasets sourced

from the OGD sources across domains for value creation and innovation (Masoumi, Farahani & Shams Aliee, 2022), but there are structural and naming conflicts that result from the differences in the URIs and the usage of different practices of modelling the structure of data cubes (Kalampokis, Karamanou & Tarabanis, 2019). Once more, in the context of Linked OGD pertaining to the Infrastructure for Spatial Information in the European Community (INSPIRE), semantic interoperability has been confronted as a potential challenge given the fact that the "INSPIRE directive doesn't specify how a resource provider could deploy services in a consistent way and how those services might better integrate spatial information with other forms of information" (Perego et al., 2012, 37).

2.2. Literature pertaining to OGD focused on the GCC region

Extant research focused on the OGD initiatives of the GCC region underline two facets clearly: OGD initiatives in the GCC region are not spearheading with the desired pace, and OGD quality in the national OGD portals of the GCC region is abysmal (Al-Sukhayri et al., 2020; Saxena, 2017a; Saxena, 2017b; Sayogo, Pardo & Cook, 2014; Tamimi, Hoshang & Al-Blooshi, 2017). Both these facets result in low citizen engagement (Al-Kubaisi, 2014). Furthermore, the governments of the GCC region are not forthcoming in terms of the implementation of the OGD initiatives or impressing upon the stakeholders the benefits that may be reaped from the re-use of OGD (Al-Rushaid & Saudagar, 2016). The last-mentioned pointer may also be the result of the lack of strategy and vision of the political leadership (Al-Kubaisi, 2018) apart from the collectivist culture being represented by the GCC regions (Saxena, 2018) given that the collectivist cultures are averse to risk-taking behaviours and attempting novelty in terms of innovations owing to their insularity.

2.3 Technical challenges with respect to OGD interoperability

Semantic interoperability is defined as "the definition of content, and deals with the human rather than machine interpretation of this content [...] interoperability at this level denotes that a common understanding exists between people regarding the definition of the content (information) being exchanged" (Rezaei et al., 2014: 6). With a focus on semantic interoperability of the OGD, the present study was designed keeping in view its accrued advantages-case in point being the OpenArchaeo for attaining the semantic interoperability by facilitating the querying of archaeological OGD on the Linked Open Data Cloud provisioned via the Tours' Laboratoired'Areologie et Territoires (France) (Marlet et al., 2019).

In general, OGD interoperability encounters technical challenges in terms of the incoherent datasets, incompleteness, repetition, uneven gap between the years and entities, lack of relevant metadata, and incompatible file formats. Furthermore, the data portal does not provide the requisite data search and filtering options alongside the fact that the tags and search strings are insufficient to cover all the datasets. Furthermore, the entry of the same dataset with a slight variation in terms of the added entities across the axes adds to the volume of the datasets to

be analysed but the actual quality assessment gets adversely impacted. Therefore, while conducting the required analyses, the precision in the similarity reports is slightly affected. Finally, and, in specific relation to the OGD focused on the UN SDGs, the insufficiency of the relevant OGD across the portals results in an incomplete analysis. This bottleneck was witnessed in the present study as well.

2.4. Research Gap

A close understanding of the aforementioned three threads (sections 2.1, 2.2 and 2.3) shows that the dimension of interoperability in the context of OGD initiatives of the GCC region has not been conducted with a focus on value creation and innovation which are the sine qua non of SDGs as well. Thus, the present study seeks to be emphatic upon the semantic interoperability of the national OGD portals of the GCC given that interoperability holds the key towards value creation and innovation. It may be pertinent to note that line of approach adopted in the present study is not new, but it has been probed elsewhere. A way in which value creation and innovation derived from OGD is possible via interoperability is showcased in terms of the ontological facet wherein the OBDM (Ontology Based Data Management) Information system identifies six dimensions that are the prerequisites for the OGD interoperability: access to the ontology documentation, analysis of the ontology, analysis of the mappings, check of the data quality, querying the system, and tuning the system (Daraio et al., 2016).

3. Methodology

In this Section, the basis of the methodological approach for this study is presented, while the steps and actions which were followed are explained in further detail. To a big extent, the present study adapts the empirical model proposed by Colpaert et al. (2014) wherein semantic interoperability is being assessed in terms of the following three metrics: "identifier interoperability, the real-world relevance of consolidated identifiers, and the number of conflicts between identifiers" (p. 50). In the context of this study, the first metric is represented by certain attributes of the examined datasets such as "title" and "description", the second metric is being approached by including datasets related to the Sustainable Development Goals (SDG's) of the examined countries, while the third metric is related to the similarity check performed on those datasets. The rationale behind this study is based on the aforementioned empirical model, however, it follows an adapted methodological approach of implementation. In the same vein, further research is warranted to measure the interoperability dimensions in the smart cities of the GCC via their OGD portals and similar (Maheshwari & Janssen, 2014).

As mentioned previously, the regional focus group of this study is the Gulf Cooperation Council (GCC). Various datasets arranged according to each one of the UN's 17 SDGs were retrieved from the national data portal of each country to be included in the analysis-this phase included a meticulous scan of the national OGD portals of the six countries and matching the specific OGD with the SDG's overarching theme and objective. Three authors were involved in

this phase and the other two authors were involved in the final review and selection process (Chokki, Alexopoulos, Saxena, Frénay, Vanderose, & Ali, 2023).

As a next step, each dataset was analyzed against several properties, including the dataset title, description, file type, language, version type and available metadata. In general, these properties can be of technical (access endpoints, protocols, standards, back-end languages, database management, and portal infrastructure) or semantic nature (catalogue metadata standards, ontologies, metadata schemata), however, in this case, special emphasis was put on the properties which would facilitate the identification and measurement of the similarity or dissimilarity among the examined datasets (Alromaih, Albassam, & Al-Khalifa, 2023; Ali, Alexopoulos, & Charalabidis, 2022). Therefore, in this light, information such as the description of the dataset content, the flat metadata and the title of the dataset were prioritized.

The obtained information was then integrated in one single file, which included all the datasets and their respective properties, as well as the 17 SDGs per GCC country. Preceding the data analysis per se, the data was prepared applying the conventional techniques for data cleaning such as handling the missing data values, case sensitivity and other structural errors. Having undergone the necessary pre-processing stage and been converted to vectors containing distinct words in the form of lists, the data was analyzed using the Python libraries, and, more specifically, the cosine similarity for textual data where the data is represented in the n-dimensional space.[1]

The cosine similarity was implemented manually using Python and some supplementary libraries such as math, pandas, regular expressions (for pre-processing the text data), and Numpy (Dang et al., 2020; Gupta et al., 2020). The cosine similarity is a measurement of proximity (cosine of the angle) between two vectors/sequences, and, typically, the outcome of the comparison is bounded for values included in the [0, 1] space (Salton, 1989). It is based on the normalized dot product of two vectors with the Euclidean normalization. The calculation of the cosine similarity formed between the two textual documents under comparison indicates their similarity; the higher the common text or context within two text documents (sentences), the higher will be the value of the cosine similarity, and vice versa. The cosine similarity takes into account the context of the textual data (Soyusiawaty & Zakaria, 2018). The formula for the calculation of the cosine similarity in text analysis is shown in Figure 1. In a nutshell, the formula checks the common words in both documents in the numerator and divides them by the sum of the unique words in both documents in the denominator.

$$similarity(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Figure 1. The calculation of the cosine similarity between two vectors, A and B.

1. The source code for the performed experiment can be found here: https://colab.research.google.com/drive/1PQ-nW1xBCmDa2T9nHxp6s948wNfKZCGZ

In the specific context of this study, each dataset vector belonging to each one of the 17 SDGs was compared against every dataset vector of **a)** the other datasets of the same SDG of the same country to measure their similarity, **b)** the other datasets of the same SDG of the other countries, so both an "intra-" and "inter-" country comparison of the datasets of each SDG is performed. Their similarity was then measured following the aforementioned method in order to, consequently, quantify and measure the semantic interoperability among the examined datasets. The result of the comparison is indicative of the distance or similarity of the compared pair. For instance, given two datasets (Dataset1 and Dataset2), the cosine similarity was calculated using the formula (Figure 1) such that the higher the Cosine similarity, the higher is the extent of interoperability, where A and B are 2 OGD metadata vectors, i is used to traverse vector A and B, and n is the total number of words which are being compared for the cosine similarity. In this case the Cosine similarity would be 0.94:

- Dataset1: Statistics on the achievements of the Supreme Council for the Environment 2019
- Dataset2: Statistics on the achievements of the Supreme Council for the Environment 2018

This example shows that the calculation of the cosine similarity for text1 and text2 was equal to 0.94, indicating a very high similarity, which is rational considering they are identical apart from the difference in the year number.

Overall, a total of 1,590 datasets were analyzed in the case of this study. Furthermore, following the described similarity check, a total of 153,515 comparisons of inter- and intra-country comparisons for the calculation of similarity indices for all the SDGs were performed. The results of the comparisons were then sorted according to their similarity index and only the datasets whose similarity index with the other datasets from different countries within the same SDG was greater than 50% (or 0.5) were considered. The rationale for establishing this threshold is to ensure that datasets with a higher degree of similarity are prioritized for interoperability. This threshold is inspired by the well-defined dice problem in statistical theory, where the probability of an event occurring is equally distributed between two possible outcomes, each with a 50% chance. In this context, success is defined as an event with a probability exceeding 50% (0.5). The described methodology was adopted with the purpose of obtaining a better understanding of the semantic alignment of the datasets, which, in its turn, provides useful insight on the level of interoperability and availability of the examined data, thus contributing to value creation. In the following Section, the Results of the data analysis are presented and interpreted for further insight.

4. Results

Before presenting the findings deriving from the analysis performed on the datasets, it is initially worthwhile to get a glimpse of the numbers describing the general status of the data involved.

Figure 2 represents the open datasets availability per SDG for each one of the examined GCC countries. It is noticeable that Saudi Arabia offers the majority of open data in this case, with a total of 1153 datasets, while UAE comes second with a total of 220 datasets for all SDGs. Qatar, Oman, Bahrain, and Kuwait follow significantly later, in this order. As far as the SDGs are concerned, the SDG with the highest availability on datasets appears to be SDG9 (Industry, Innovation, and Infrastructure), followed by SDG3 (Good Health and Well-being), and SDG4 (Quality Education).

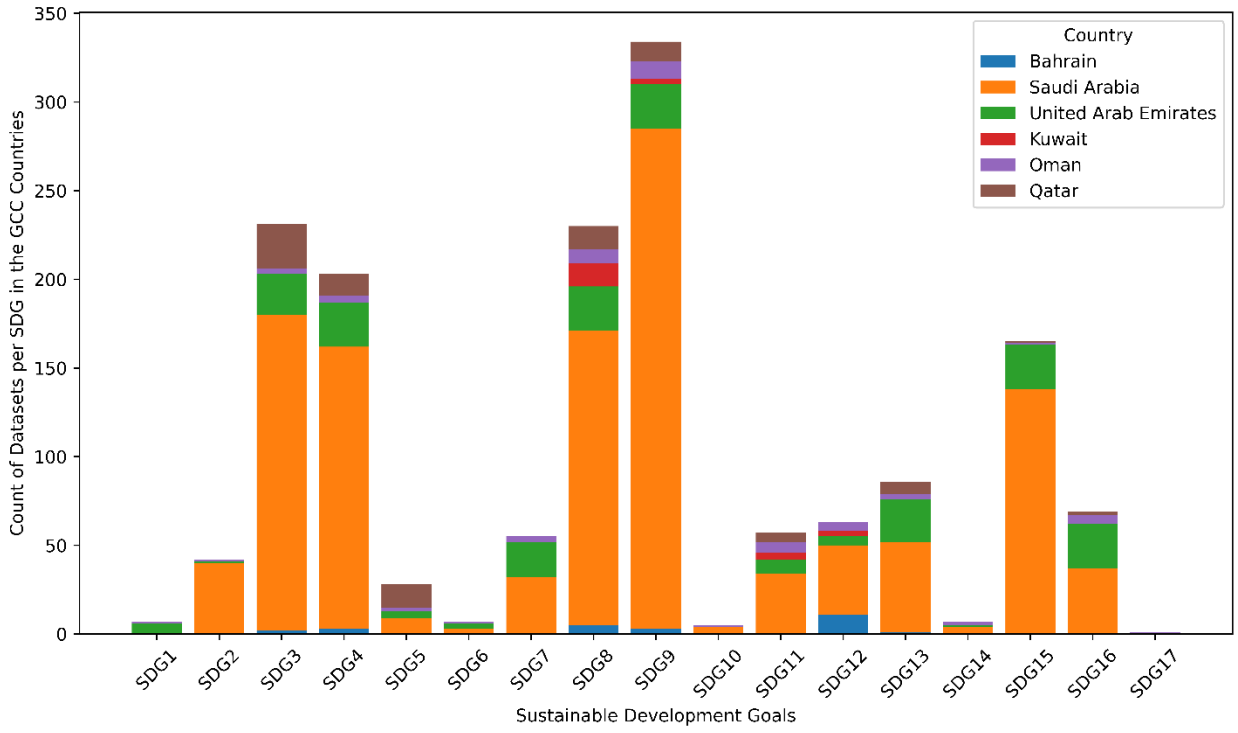

Figure 2. Open datasets availability per SDGs in GCC countries

Saudi Arabia and the United Arab Emirates are significantly contributing more data pertinent to the Sustainable Development Goals (SDGs), accounting for more than 75% of the entire dataset collected. This study examines the availability of Open Government Data (OGD) across the Gulf Cooperation Council (GCC) countries and evaluates how sufficient and interoperable this OGD is. One more thing is worth mentioning: individually, GCC countries do not have the requisite OGD data to quantify interoperability; however, cumulatively, interoperability quantification is realisable.

In order to better visualize the proportion of datasets and open data contribution per country, the pie chart of Figure 3 provides the percentages of data availability in the GCC countries. Saudi Arabia provides a maximum proportion of datasets followed by the United Arab Emirates, Qatar, Oman, and then equal contribution by Kuwait and Bahrain follow.

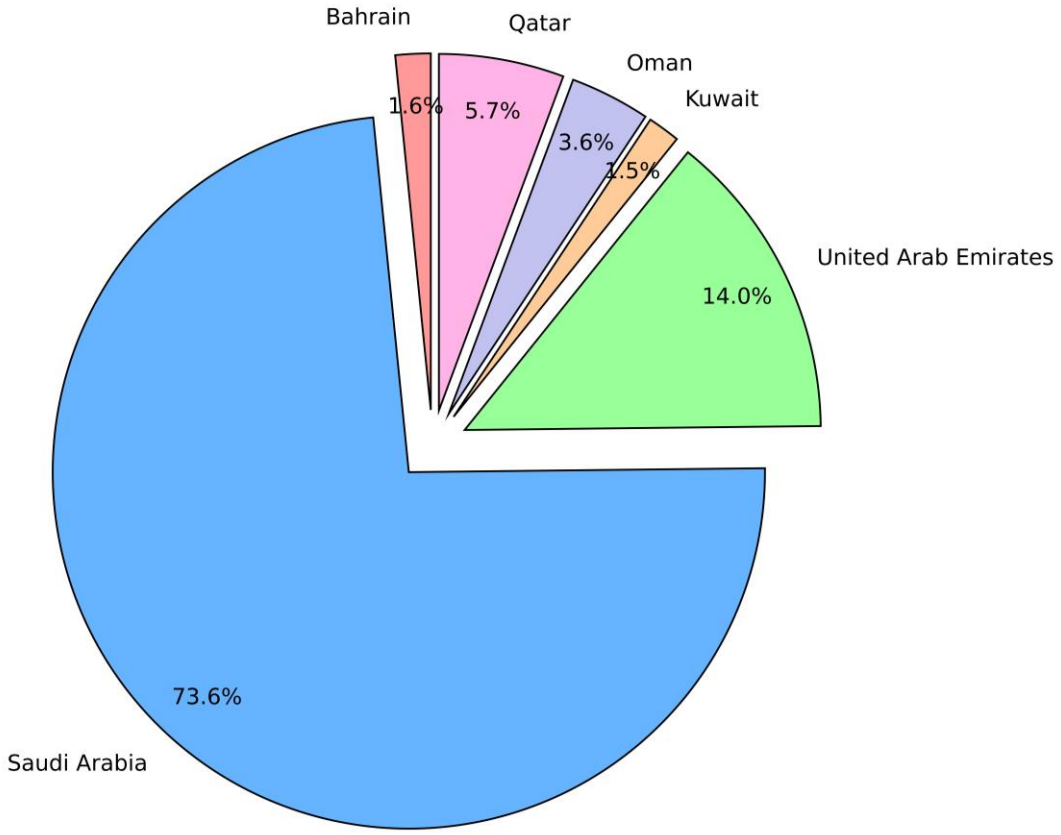

Figure 3. Open data availability for SDGs in GCC countries

As mentioned previously, each dataset vector of the 17 SDGs was compared against every dataset vector of **a)** the other datasets of the same SDG of the same country to measure their similarity, **b)** the other datasets of the same SDG of the other countries. The similarity indices were calculated for every SDG and a total of 153,515 comparisons of inter- and intra-country comparisons of the dataset vectors took place. As an output of this process, and to capture the essence of a feasible visualization, the average similarity index of the open datasets for each SDG is presented in Figure 4.

The similarity index is indicative of the interoperability of the datasets, since the process involved the comparison of properties of the dataset (title, description, file type, version type, language, and metadata) which, if highly similar, act as a facilitator for shared meaning (semantic interoperability). In this case (and as can be seen in Figure 4), only the interoperability values higher than 0.5 were considered. In other words, Figure 4 can be viewed as a representation of the most interoperable SDGs examined, which in this case, are SDG8 (Decent work and economic growth), SDG 13 (Climate Action) and SDG 9 (Industry, Innovation and Infrastructure) with values of 0.68, 0.68, and 0.66 respectively. These are followed up by SDG4 (Quality Education), SDG12 (Responsible Consumption and Production), SDG3 (Good Health and Well-being) and SDG 11(Sustainable Cities and Communities) in that order. Figure 4 helps in determining the contributions of each country towards the interoperability of datasets in SDGs.

The countries that are playing a role in each SDG's interoperability are also mentioned with the SDG number on the x-axis. SDGs without any value represents that no country is playing a role in this SDG for interoperability. The SDG1, SG6, and SDG17 do not have any countries or datasets that are interoperable, which is the reason the authors have indicated them as "SDG#: without any vertical bar." For the remaining SDGs and their interoperability, the SDG numbers with the GCC countries' abbreviations have been denoted on the x-axis. For instance, in SDG3, the cumulative similarity index is 0.65, and countries in which SDG3 is interoperable are the United Arab Emirates, Qatar, Bahrain, and Saudi Arabia. Similarly, SDG4 has a cumulative similarity index of 0.69, and countries involved in this interoperability of SDG4 are Qatar, the United Arab Emirates, Saudi Arabia and Oman. In this way, the remaining SDGs have also been explained.

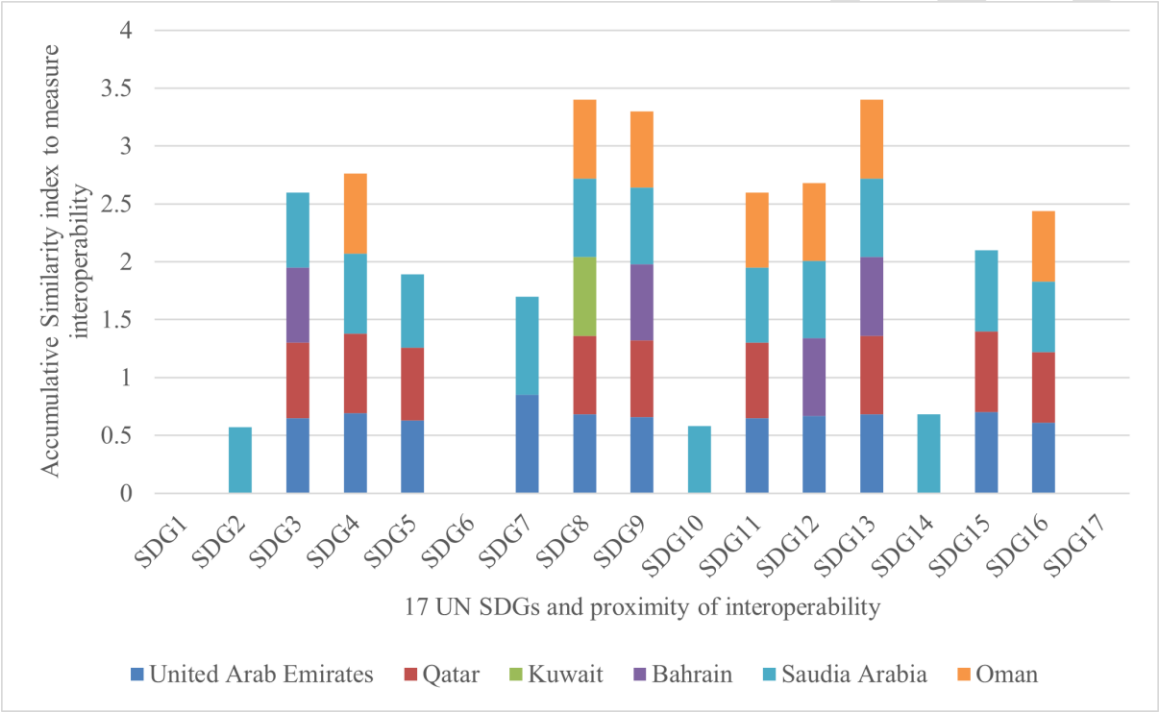

Figure 4. The similarity index as interoperability measure of the open datasets for the 17 SDGs in GCC countries

Figure 4 illustrates the cumulative similarity index and the contributions of individual GCC countries towards these similarity indices, serving as a quantification of the interoperability of the datasets within each SDG. It's also evident that the values of the cumulative similarity index, ranging from 0 to 1, for each category (SDGs) and their corresponding segments (countries), contribute to the interoperability of datasets. During this experimental phase, we aim to compare the interoperability and availability of datasets that contribute to the SDGs. Comparing interoperability and availability requires the same scale and range. As already elaborated, interoperability is in the range [0, 1]. However, we need to assure the availability of datasets that align with the interoperability range of [0, 1] (Al Shalabi & Shaaban, 2006; Kiran &

Vasumathi, 2020). To accomplish this, Min-max normalization for standard range [0, 1] technique was applied; the formula is shown in Eq. 1.

$$Normalized\ Availability = \frac{Origional\ Availability(x) - Minimum\ value\ of\ the\ availability}{Maximum\ value\ of\ the\ availability - Minimum\ value\ of\ the\ availability} \quad (1)$$

Figure 5 presents a scatter plot that simultaneously displays the normalized values for the availability and interoperability of the SDGs, illustrating their relationship to each other. The availability of Sustainable Development Goals (SDGs) datasets, as depicted in Figure 2, exhibits significant variations among countries. For instance, Saudi Arabia surpasses other nations in the provision of datasets. However, when examining interoperability, as illustrated in Figure 5, it becomes evident that there exists a uniform level of interoperability potential within the Gulf Cooperation Council (GCC) countries. To elucidate, if we take SDG3 as an example, it is noteworthy that four countries—namely the United Arab Emirates, Qatar, Bahrain, and Saudi Arabia—are identified as participants. This implies that these nations share an equitable opportunity for open data interoperability, irrespective of variations in data availability, owing to the presence of the required datasets within one of the partnering GCC countries. Figure 5 smooth the way to compare the availability and interoperability of datasets pertaining to SDGs.

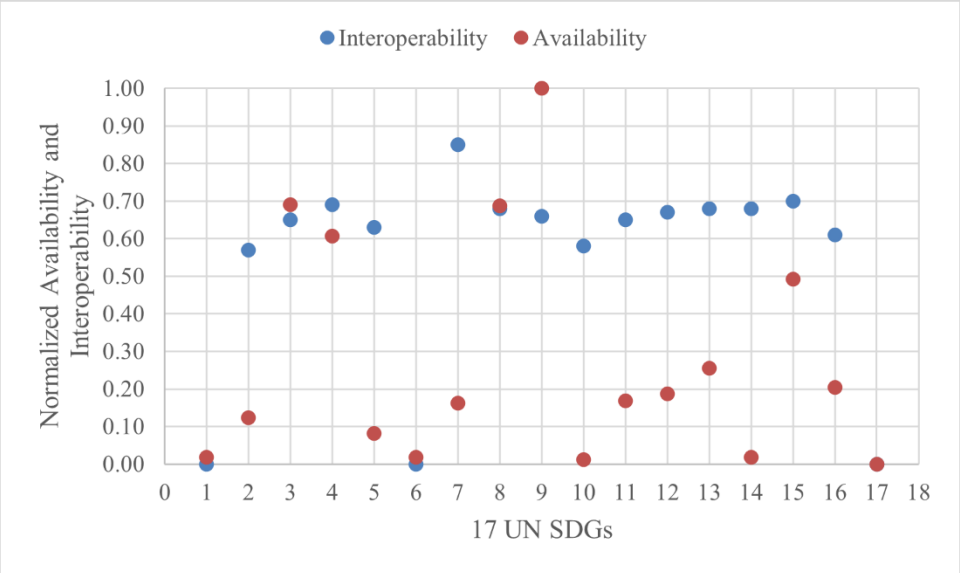

Figure 5. Scatter plot of the normalized values for availability and interoperability

From this scatter plot shown in Figure 5, the authors extracted the quadrant visualization for the availability and interoperability of open datasets for the 17 SDGs in the GCC countries. The Availability-Interoperability quadrant is shown in Figure 6.

The graph quadrant typically represents the four sections on a Cartesian plane (discernibly, only positive values in this case) and is showing the distribution of the 17 examined SDGs, according to their score for availability and interoperability. In line with their quadrant

positioning, the following results may be drawn:

I. *Low Availability-Low Interoperability*: SDG 1 (No poverty); SDG 6 (Clean Water and Sanitation), SDG 17 (Partnerships for the goals).

II. *Low Availability-High Interoperability*: SDG 2 (Zero Hunger), SDG 5 (Gender Equality), SDG 7 (Affordable and Clean energy), SDG 10 (Reduced Inequalities), SDG 11 (Sustainable Cities and Communities), SDG 12 (Responsible Consumption and Production), SDG 13 (Climate Action), SDG 14 (Life below water), SDG 16 (Peace, justice and strong institutions).

III. *High Availability-Low Interoperability*: None.

IV. *High Availability-High Interoperability*: SDG 3 (Good health and well-being), SDG 4 (Quality Education), SDG 8 (Decent Work and Economic Growth), SDG 9 (Industry, Innovation and Infrastructure), SDG 15 (Life on Land).

Apart from the cases where the low/high availability and low/high interoperability are matched and rightly so, the confounding results for the complete absence of high availability-low interoperability SDG-related datasets are difficult to comprehend. We cannot infer from the graph quadrant whether the availability of OGD and their interoperability are strongly interrelated. The absence of instances in the quadrant of high_availability-low_interoperability could only be an indication of this interrelation regarding SDGs datasets.

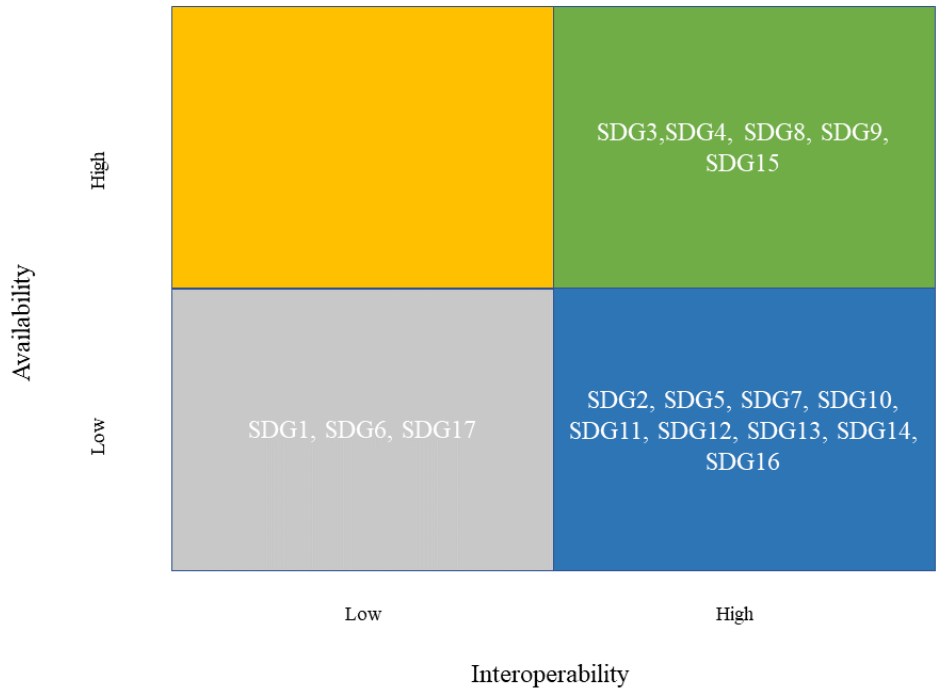

Figure 6. The Availability - Interoperability Quadrant for the 17 SDGs (Source: Authors)

On the other hand, mere availability of OGD is not a surety of being qualitatively compatible (Janssen, Charalabidis & Zuiderwijk, 2012). Also, mere availability of OGD is not a surety for

its quality or being linked to SDG, for that matter and this reduces the chances of interoperability as well. Likewise, the SDGs falling in the quadrant pertaining to the low availability-high interoperability SDGs datasets are equally mismatched given that the interoperability between minimal number of datasets does not warrant qualitative adequacy, at least, because interoperability will lose its relevance with inadequate sample size (Wilkinson et al., 2016). This last dimension may be a result of the lesser degree of contribution from the government departments and/or the lesser prioritized sectors covered by them, which are nevertheless interoperable in terms of their similarity indices.

5. Discussion

Interoperability, and hence the value creation and innovation drives, is a challenge in the developing as well as the developed countries alike-case in point being the interoperability issues pertaining to the health-related OGD in the African region (Oluoch et al., 2021). As such, the GCC region was poised to grow in terms of its economic growth by more than 6.9% in 2022 (World Bank, 2023) and it is anticipated that the OGD initiatives of the GCC region might be instrumental in providing robust SDG-related solutions via the value creation and innovation activities made possible by the interoperability of the OGD. Whilst some of the results were consequential and indicative of the applicative character of the SDGs via the OGD availability (high availability-high interoperability) (Williams et al., 2012), the loose spatial interconnectedness may be a significant reason for the imperfect match between the availability-interoperability quadrant analysis (Mohammadi, Rajabifard & Williamson, 2010). Thus, the metadata and the actual data might be incoherent (Wiemann & Bernard, 2016) in some cases as well which explains the findings pertaining to high availability-low interoperability graph quadrant. Another potential reason for the last-mentioned finding may be owing to the lack of a common understanding among the different stakeholders or their complete exclusion from the OGD publishing processes-case in point being the EU Integrated Project 507849 ATHENA (Advanced Technologies for Interoperability of Heterogeneous Enterprise Networks and their Application) (Petersen, Lillehagen & Anastasiou, 2008).

Furthermore, the case of high availability-low interoperability may be owing to the differentials in timeframe of OGD publishing such that the cases whereas the quantitatively rich historic data is readily available, but the requisite real-time data is not readily available or lacks granularity, there is a consequent interoperability (Douthit et al., 2021). This instance of high availability-low interoperability was also evident in the datasets pertaining to the disaster risk governance managed by the European Science and Technology Advisory Group (Migliorini et al., 2019) as also the datasets maintained by the Kadaster, the Dutch National Land Registry and Mapping Agency (Rowland et al., 2022).

Regarding the OGD availability vis-a-vis interoperability pertaining to attaining the SDGs' achievement, the findings are in line with our expectations. For instance, regarding the high availability-high interoperability instances, it may be comprehended that GCC is a business hub (Jouini, 2015) with impetus on Smart Living (Samad & Azar, 2019) and Quality Education (Al-Sharari, 2018) which explains the quadrant characteristics. Furthermore, the low availability-

high interoperability instances, the related SDGs for climate change or energy, for instance, are few on account of the predictable nature of climate (Al-Maamary, Kazem & Chaichan, 2017) and the very recent shift towards clean energy (Alharbi & Csala, 2020; Saqib, 2018) on account of the depleting hydrocarbon-based resources which had led to scarcity of OGD pertaining to clean and efficient energy.

As far as the high availability-low interoperability absence of instances is concerned, it may be pointed out that the relevant SDG-oriented OGD are conspicuously missing or perhaps of poor quality or missing values and/or metadata or lacking machine-processability and this factoid has been attested in prior research (Alromaih, Albassam & Al-Khalifa, 2016; Mutambik et al., 2022). Finally, owing to low- or poor-quality SDG-focused OGD, the interoperability possibilities diminish accordingly. The exclusive unique datasets in the GCC portals are also one reason behind the high availability-low interoperability datasets. GCC countries need to put effort into diverging the “highly available and low-interoperable” datasets, if they have them in the future, to convert them into "highly available and high-interoperable" datasets.

Regarding the low availability-low interoperability instances, it is a logical conclusion, but it is surprising that the OGD pertaining to SDG 1, 2, 6, 10, 17 are scanty despite being so relevant. It is understandable that SDG 14 related OGD is less on account of the minimal number of water bodies in the region. The results compiled in this study explain how availability and interoperability are interlinked to each other. For instance, Saudi Arabian, UAE and Qatar have multifarious SDGs-related open datasets availability and that's why these datasets are significantly interoperable than the other country’s datasets like Oman and Kuwait. Consequently, this is a win-win situation for the availability and interoperability of the datasets, the higher the availability, the higher the interoperability potential of the datasets. However, high-availability does not explicitly indicate high-interoperability in all cases for the reasons cited above.

To summarize, the greater the availability and interoperability of the SDG-related open datasets, the more they will play a significant role in value-creation and the attainment of sustainable development goals in GCC countries.

6. Academic implications

The present study leaves significant research directions. Further research is warranted to assess the extent to which the national OGD portals of the GCC facilitate the interoperability of metadata in the linked OGD environment (Sugimoto et al., 2016). Likewise, interoperability of OGD with a specific focus on the SDGs may be investigated from the perspectives of the users themselves as to the facilitating and hindering determinants for value creation and innovation. Since the present study focused on the GCC countries, it is anticipated that further research would underscore the benchmarks for realizing the objectives appropriately by drawing a comparative perspective vis-a-vis the developed countries’ OGD initiatives. Another open issue is the efficiency improvement of the measuring algorithm. The time complexity of the similarity measuring algorithm for measuring the interoperability of open datasets per SDG is excessively

high, because it must compare each dataset within an SDG with other datasets within the same SDG. The increase in countries and in datasets per Sustainable Development Goal will also increase the time complexity of similarity measurements between these datasets. This will lead to the parallelization of the developed algorithm.

7. Practitioner implications

The present study leaves insights for the practitioners as well. For one, it is important that the knowledge management happens with the interoperability applications while focusing on the attributive dimensions of the OGD (da Costa Castro et al., 2017). The main emphasis of the government agencies should be to ensure that the data availability is robust and qualitatively advanced with the requisite infrastructure apart from collaborative arrangements with the concerned stakeholders. While conducting the present study, researchers encountered bottlenecks in terms of the technicalities entailed while searching and filtering the required OGD-case in point being the repetitive OGD and the incoherent formats and structures of the data and source files. Furthermore, there were issues related to the insufficient metadata accompanying the OGD. Therefore, the policy makers should spearhead their efforts towards improvising the OGD initiatives. Policy makers need to align their OGD initiatives towards the realization of the SDGs via easily accessible and statistically operative OGD to ensure the institutionalization and sustainability of the OGD initiatives. Their strategies could be partially driven by the two major factors in the quadrant. Apart from strategic vision and planning for the realisation of SDGs via value creation initiatives from the stakeholders concerned, the bottom line is to align the OGD provisioning in tandem with the SDGs and provide high-value SDG-focused OGD to facilitate interoperability. This is possible only with the commitment and engagement of the concerned stakeholders-both from the side of the suppliers, i.e., the government, and from the side of the users, i.e., the citizens, app developers, entrepreneurs, voluntary sector and so on.

8. Conclusion

The focus of the study was to underline the interoperability possibilities leading to value creation and innovation in order to achieve the SDGs via the OGD available across the portals of the GCC. Interoperability of datasets has been found to have significant implications for the diverse socio-economic sectors like health (Kouroubali & Katehakis, 2019), agriculture (Porter et al., 2014), ecosystem (Buck et al., 2019) and the like. With this background, an empirical investigation protocol was adopted wherein the cosine similarity measures were assessed across the SDG-relevant OGDs of six GCC countries to understand the possibilities of interoperability for driving forth value creation and innovation by the stakeholders. Findings from the study show that the GCC countries have provisioned heterogenous OGD across diverse socio-economic sectors but there is no guarantee of them being interoperable in line with the attainment of the SDGs via value creation and innovation activities. The present study also shows that semantic interoperability of OGD gets hindered on account of the differences in

scope, terminology, metadata fields, quality and structure of the data descriptions (Bouchalouche, Ghomari & Zemmouchi-Ghomari, 2022; Maali, Cyganiak & Peristeras, 2010).

The mismatch between the OGD availability and interoperability feasibility, for instance regarding the OGD pertaining to energy or water and sanitation, poses a challenge before the stakeholders to find sustainable solutions for addressing the SDGs' requirements. At the same time, there are instances where the SDGs-relevant OGD is available and there is increased potential of their being interoperable, case in point being those related with businesses or innovation and infrastructure. Finally, the analysis and results provide a tool to the policy makers to further advance their OGD strategies regarding SDGs. The final quadrant analysis along with figures 2, 3 and 4 would drive the strategic development for each individual country to move from the acquired level to the next one.

This study comes with some limitations. To begin with, the number of datasets per SDG that was unequal so the semantic interoperability assessment could not be performed. In addition, the successful and unsuccessful case studies of value creation and innovation via the SDGs have not been investigated to validate the results, and the cosine similarity assesses direction but not the magnitude (Zhou et al., 2022). Nonetheless, as a first contribution towards the OGD-literature regarding the possibilities of value creation via the SDG-related OGD, the present study is a firm grounding for furthering our understanding of the efficacious functioning of the OGD ecosystem balancing the value creation and innovation endeavours, while at the same time contributing to the attainment of SDG goals. For future research, this study could be extended by utilizing advanced language models, such as large language models (LLMs) or generative AI, to analyse not only metadata but also big, open, and linked datasets directly. Moreover, extending this research to the European Open Data portal, which features around 1.5 million datasets from European member states, would greatly enhance the assessment and implementation of the SDGs and foster innovation and value creation in both the public and private sectors.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955569. The opinions expressed in this document reflect only the author's view and in no way reflect the European Commission's opinions. The European Commission is not responsible for any use that may be made of the information it contains.

Conflict of Interest

Authors have no conflict of interest whatsoever.

References

Ahlgren, B., Hidell, M., Ngai, E.C.H. Internet of Things for smart cities: Interoperability and

open data. IEEE Internet Computing. **20**(6), 52-56. (2016). https://doi.org/10.1109/MIC.2016.124.

Alexopoulos, C., Spiliotopoulou, L., Charalabidis, Y. Open data movement in Greece: A case study on open government sources. Proceedings of the 17th Panhellenic Conference on Informatics. 279-286. (2013). https://doi.org/10.1145/2491845.2491876.

Alharbi, F.R. Csala, D. GCC countries' renewable energy penetration and the progress of their energy sector projects. IEEE Access. **8**, 211986-212002. (2020). https://doi.org/10.1109/ACCESS.2020.3039936.

Al-Kubaisi, A.S. Improving the transparency, openness and efficiency of e-government in Qatar in the era of Open Government Data, and beyond. PhD thesis submitted to the Faculty of Law, Queensland University of Technology. (2014). Available at https://eprints.qut.edu.au/75540/2/Ali_Al-Kubaisi_Thesis.pdf, accessed on 20th January, 2022.

Al-Kubaisi, A.S. Enhancing the adoption of e-government systems through Open Government and Open Government Data (OGD) initiatives in Qatar. Journal of Information Engineering and Applications. **8**(1) (2018). https://www.iiste.org/Journals/index.php/JIEA/article/view/40706.

Al-Maamary, H.M.S., Kazem, H.A., Chaichan, M.T. Climate change: The game changer in the gulf cooperation council region. Renewable and Sustainable Energy Reviews. **76**, 555-576. (2017). https://doi.org/10.1016/j.rser.2017.03.048.

Alromaih, N., Albassam, H., Al-Khalifa, H. A proposed checklist for the technical maturity of open government data: An application on GCC countries. 18th International Conference on Information Integration and Web-based Applications and Services, 494-499. (2016). https://doi.org/10.1145/3011141.3011211.

Al Rushaid, M.W., Saudagar, A.K.J. Measuring the data openness for the Open Data in Saudi Arabia e-government-A case study. International Journal of Advanced Computer Science and Applications. **7**(12), 113-122. (2016). https://dx.doi.org/10.14569/IJACSA.2016.071215.

Al Shalabi, L., Shaaban, Z. Normalization as a preprocessing engine for data mining and the approach of preference matrix. International Conference on Dependability of Computer Systems, SzklarskaPoreba, Poland, 207-214. (2006).https://doi.org/10.1109/DEPCOS-RELCOMEX.2006.38.

Alsharari, N.M. Internationalization of the higher education system: An interpretive analysis. International Journal of Educational Management. **32**(3), 359-381. (2018).

https://doi.org/10.1108/IJEM-04-2017-0082.

Al Sukhayri, A.M., Aslam, M.A., Saeedi, K., Malik, M.S.A. A linked open data-oriented sustainable system for transparency and open access to government data: A case study of the public's response to women's driving in Saudi Arabia. Sustainability. **12**(20), 8608. (2020). https://doi.org/10.3390/su12208608.

Bertot, J.C., Gorham, U., Jaeger, P.T., Sarin, L.C., Choi, H. Big data, open government and e-government: Issues, policies and recommendations. Information Polity. **19**(1-2), 5-16. (2014). https://doi.org/10.1145/2479724.2479730.

Bouchelouche, K., Ghomari, A.L. Zemmouchi-Ghomari, L. Enhanced analysis of Open Government Data: Proposed metrics for improving data quality assessment. 5th International Symposium on Informatics and its Applications (ISIA), M'sila, Algeria, 1-6. (2022).https://doi.org/10.1109/ISIA55826.2022.9993482.

Buck, J.H., Bainbridge, S.J., et al. Ocean data product integration through innovation-The next level of data interoperability. *Frontiers in Marine Science*. **6**. (2019). https://doi.org/10.3389/fmars.2019.00032.

Charalabidis, Y., Alexopoulos, C., Loukis, E. A taxonomy of open government data research areas and topics. Journal of Organizational Computing and Electronic Commerce. **26**(1-2), 41-63. (2016). https://doi.org/10.1080/10919392.2015.1124720.

Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T., Ferro, E. Open Data Interoperability. In: *The World of Open Data. Public Administration and Information Technology*, 28. Springer, Cham. (2018). https://doi.org/10.1007/978-3-319-90850-2_5.

Chaturvedi, K., Kolbe, T.H. Towards establishing cross-platform interoperability for sensors in smart cities. Sensors. **19**, 562-591. (2019). https://doi.org/10.3390/s19030562.

Cho, J. Comparative analysis of open government data topics and usability. Qual Quant 57, 5655–5671 (2023). https://doi.org/10.1007/s11135-023-01630-x

Colpaert, P., Van Compernolle, M., De Vocht, L., Dimou, A., Sande, M.V., Verborgh, R., Mechant, P., Mannens, E. Quantifying the interoperability of open government datasets. Computer. **47**, 50-56. (2014). https://doi.org/https://10.1109/MC.2014.296.

da Costa Castro, A., Cestari, J.M.A., Loures, E.R., de Lima, E.P., Santos, E.A.P. Interoperability Frameworks in Public Administration Domain: Focus on Enterprise Assessment. In: Amorim, M., Ferreira, C., Vieira Junior, M., Prado, C. (eds) *Engineering Systems and Networks*. Lecture Notes in Management and Industrial Engineering. Springer, Cham. (2017).

https://doi.org/10.1007/978-3-319-45748-2_20.

Dang, L-h., Feng, N., An, Guo-s., Cao, J., Du, Q-x., Jin, Q-q., Huang, P., Sun, J-h. Novel insights into wound age estimation: combined with "up, no change, or down" system and cosine similarity in python environment. International Journal of Legal Medicine. **134**, 2177-2186. (2020). https://doi.org/10.1007/s00414-020-02411-z.

Daraio, C., Lenzerini, M., Leporelli, C., Naggar, P., Bonaccorsi, A., Barttolucci, A. The advantages of an Ontology-Based Data Management approach: openness, interoperability and data quality. Scientometrics. **108**, 441-455. (2016). https://doi.org/10.1007/s11192-016-1913-6.

De Blasio, E., Selva, D. Implementing open government: a qualitative comparative analysis of digital platforms in France, Italy and United Kingdom. Qual Quant 53, 871–896 (2019). https://doi.org/10.1007/s11135-018-0793-7

Douthit, B.J., Fiol, G.D., Staes, C.J., Docherty, S.L. Richesson, R.L. A conceptual framework of data readiness: The contextual intersection of quality, availability, interoperability, and provenance. Applied Clinical Informatics. **12**(3), 675-685. (2021). https://doi.org/10.1055/s-0041-1732423.

Gonzalvez-Gallego, N., Nieto-Torrejon, L., Perez-Carceles, M.C. Is open data an enabler for trust? Exploring the link and the mediating role of citizen satisfaction. International Journal of Public Administration. **43**(14), 1218-1227. (2020). https://doi.org/10.1080/01900692.2019.1668412.

Gupta, M., Thakkar, A., Aashish, Gupta, V. Rathore, D.P.S. Movie recommender system using collaborative filtering. International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 415-420. (2020). https://doi.org/10.1109/ICESC48915.2020.9155879.

Hernandez, J., Garcia, R., Schonowski, J., Atlan, D., Chanson, G., Ruohomaki, T. Interoperable open specifications framework for the implementation of standardized urban platforms. Sensors. **20**, 2402, (2020).https://doi.org/10.3390/s20082402.

Janowski, T. Implementing Sustainable Development Goals with digital government – Aspiration-capacity gap. Government Information Quarterly. **33**(4), 603-613. (2016). https://doi.org/10.1016/j.giq.2016.12.001.

Janssen, K. The influence of the PSI directive on open government data: An overview of recent developments. Government Information Quarterly. **28**(4), 446-456. (2011). https://doi.org/10.1016/j.giq.2011.01.004.

Janssen, M., Charalabidis, Y., Zuiderwijk, A. Benefits, adoption barriers and myths of open data

and open government. Information Systems Management. **29**(4), 258-268. (2012). https://doi.org/10.1080/10580530.2012.716740.

Jetzek, T., Avital, M., Bjorn-Andersen, N. The generative mechanisms of open government data. In: *Proceedings of the 21st European Conference on Information Systems (ECIS)*, 156, Utrecht, The Netherlands. (2013). https://aisel.aisnet.org/ecis2013_cr/156/.

Janssen, M., Estevez, E., Janowski, T. Interoperability in Big, Open, and Linked Data-Organizational maturity, capabilities, and data portfolios. Computer. **47**(10), 44-49. (2014). https://doi.org/10.1109/MC.2014.290.

Jimenez, C.E., Solanas, A., Falcone, F. E-Government Interoperability: Linking Open and Smart Government. Computer. **47**(10), 22-24. (2014). https://doi.org/10.1109/MC.2014.281.

Jouini, J. Linkage between international trade and economic growth in GCC countries: Empirical evidence from PMG estimation approach. The Journal of International Trade & Economic Development. **24**(3), 341-372. (2015). https://doi.org/10.1080/09638199.2014.904394.

Kalampokis, E., Karamanou, A., Trabanis, K. Interoperability conflicts in linked open statistical data. Information. **10**, 249. (2019). https://doi.org/10.3390/info10080249.

Kiran, A.,Vasumathi, D. Data mining: Min–max normalization based data perturbation technique for privacy preservation. In: Raju, K., Govardhan, A., Rani, B., Sridevi, R., Murty, M. (eds) *Proceedings of the Third International Conference on Computational Intelligence and Informatics*. Advances in Intelligent Systems and Computing, 1090. Springer, Singapore. (2020). https://doi.org/10.1007/978-981-15-1480-7_66.

Kouroubali, A., Katehakis, D.G. The new European interoperability framework as a facilitator of digital transformation for citizen empowerment. Journal of Biomedical Informatics. 94, 103166. (2019). https://doi.org/10.1016/j.jbi.2019.103166.

Maali, F., Cyganiak, R., Peristeras, V. Enabling Interoperability of Government Data Catalogues. In: Wimmer, M.A., Chappelet, JL., Janssen, M., Scholl, H.J. (eds) Electronic Government. EGOV 2010. Lecture Notes in Computer Science, **6228**. Springer, Berlin, Heidelberg. (2010).https://doi.org/10.1007/978-3-642-14799-9_29.

Maheshwari, D., Janssen, M. Reconceptualizing measuring, benchmarking for improving interoperability in smart ecosystems: The effect of ubiquitous data and crowdsourcing. Government Information Quarterly. **31**(1), S84-S92. (2014). https://doi.org/10.1016/j.giq.2014.01.009.

Marlet, O., Francart, T., Markhoff, B., Rodier, X. OpenArchaeo for usable semantic interoperability. ODOCH. (2019).https://hal.archives-ouvertes.fr/hal-02389929.

Maretti, M., Russo, V. & del Gobbo, E. Open data governance: civic hacking movement, topics and opinions in digital space. Qual Quant 55, 1133–1154 (2021). https://doi.org/10.1007/s11135-020-01045-y

Medaglia, R., Misuraca, G., Aquaro, V. Digital government and the United Nations' Sustainable Development Goals: Towards an analytical framework. DG.O2021: The 22nd Annual International Conference on Digital Government Research, 473-478. (2021). https://doi.org/10.1145/3463677.3463736.

Migliorini, M., Hagen, J.S., Mihaljević, J., Mysiak, J., Rossi, J.-L., Siegmund, A., Meliksetian, K., Guha Sapir, D. Data interoperability for disaster risk reduction in Europe. Disaster Prevention and Management. **28**(6), 804-816. (2019). https://doi.org/10.1108/DPM-09-2019-0291.

Mohammadi, H., Rajabifard, A., Williamson, I.P. Development of an interoperable tool to facilitate spatial data integration in the context of SDI. International Journal of Geographical Information Science. **24**(4), 487-505. (2010). https://doi.org/10.1080/13658810902881903.

Morando, F. Legal interoperability: making Open Government Data compatible with businesses and communities. JLIS.It. **4**(1), 441. (2013). https://doi.org/10.4403/jlis.it-5461.

Mutambik, I., Almuqrin, A., Lee, J., Gauthier, J., Homadi, A. Open government data in Gulf Cooperation Council countries: An analysis of progress. Sustainability. **14**(12), 7200. (2022). https://doi.org/10.3390/su14127200.

Oluoch, T., Muturi, D., Kiriinya, R., Waruru, A., Lanyo, K., et al. Do interoperable national information systems enhance availability of data to assess the effect of scale-up of HIV services on health workforce deployment in resource-limited countries? Studies in Health Technology and Informatics. **216**, 677-681. (2016). https://doi.org/10.3233/978-1-61499-564-7-677.

Perego, A. Fugazza, C., Vaccari, L., Lutz, M., Smits, P., Kanellopoulos, I., Schade, S. Harmonization and interoperability of EU environmental information and services. IEEE Intelligent Systems. **27**(3), 33-39. (2012).https://doi.org/10.1109/MIS.2012.22.

Petersen, S.A., Lillehagen, F., Anastasiou, M. Modelling and Visualisation for Interoperability Requirements Elicitation and Validation. In: Manolopoulos, Y., Filipe, J., Constantopoulos, P., Cordeiro, J. (eds) Enterprise Information Systems. ICEIS 2006. Lecture Notes in Business Information Processing. **3**. Springer, Berlin, Heidelberg. (2008). https://doi.org/10.1007/978-3-540-77581-2_16.

Lodato, T., French, E., Clark, J. Open government data in the smart city: Interoperability, urban knowledge, and linking legacy systems. Journal of Urban Affairs. **43**(4), 586-600. (2021). https://doi.org/10.1080/07352166.2018.1511798.

Masoumi, H., Farahani, B., Shams Aliee, F. Systematic and ontology-based approach to interoperable cross-domain open government data services. Transforming Government: People, Process and Policy. **16**(1), 110-127. (2022). https://doi.org/10.1108/TG-08-2021-0132.

Porter, C.H., Villalobos, C., Holzworth, D., Nelson, R. et al. Harmonization and translation of crop modeling data to ensure interoperability. Environmental Modelling & Software. **62**, 495-508. (2014). http://dx.doi.org/10.1016/j.envsoft.2014.09.004.

Rezaei, R., Chiew, T.K., Lee, S.P. A review on e-business interoperability frameworks. Journal of Systems and Software. **93**, 199-216. (2014).https://doi.org/10.1016/j.jss.2014.02.004.

Rowland, A., Folmer, E., Beek, E., Wenneker, R. Interoperability and integration: An updated approach to linked data publication at the Dutch Land Registry. ISPRS International Journal of Geo-Information. **11**(1), 51, (2022). https://doi.org/10.3390/ijgi11010051.

Salton, G. *Automatic text processing: The transformation, analysis, and retrieval of information by computer*. Addison-Wesley Longman Publishing Co. Inc., MA, US. (1989).

Samad, W.A., Azar, E. *Smart cities in the Gulf: An overview*. In: Samad, W.A., Azar, E. (eds) Smart Cities in the Gulf. Palgrave Macmillan, Singapore. (2019). https://doi.org/10.1007/978-981-13-2011-8_1.

Saqib, N. Greenhouse gas emissions, energy consumption and economic growth: Empirical evidence from gulf cooperation council countries. International Journal of Energy Economics and Policy. **8**(6), 392-400. (2018). https://www.econjournals.com/index.php/ijeep/article/view/7269.

Saxena, S. Open public data (OPD) and the Gulf Cooperation Council (GCC): Challenges and prospects. Contemporary Arab Affairs. **10**(2), 228-240. (2017a). https://doi.org/10.1080/17550912.2017.1297565.

Saxena, S. Significance of open government data in the GCC countries. Digital Policy, Regulation and Governance. **19**(3), 251-263. (2017b). https://doi.org/10.1108/DPRG-02-2017-0005.

Saxena, S. National open data frames across Japan, The Netherlands and Saudi Arabia: Role of culture. Foresight. **20**(1), 123-134. (2018). https://doi.org/10.1108/FS-07-2017-0038.

Sayogo, D.S., Pardo, T.A., Cook, M. A framework for benchmarking Open Government Data efforts. 47th Hawaii International Conference on System Sciences, 1896-1905, (2014). Available at https://ieeexplore.ieee.org/abstract/document/6758838, accessed on 4 December, 2022.

Soyusiawaty, D., Zakaria, Y. Book data content similarity detector with cosine similarity (Case study on digilib.uad.ac.id). *12th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, Yogyakarta, Indonesia, 1-6. (2018). https://doi.org/10.1109/TSSA.2018.8708758.

Sugimoto, S., Li, C., Nagamori, M., Greenberg, J. (2016). Permanence and temporal interoperability of metadata in the linked open data environment. *Proceedings of the International Conference on Dublin Core and Metadata Applications*, 45-54.https://dl.acm.org/doi/10.5555/3183795.3183816.

Tamimi, H., Hoshang, S.A., Al Blooshi, E.J. Analysis of UAE open government data usability within mobile application development. IEEE 2nd International Conference on Big Data Analysis (ICBDA), 437-441. (2017). https://doi.org/10.1109/ICBDA.2017.8078857.

UN. *About the Sustainable Development Goals. United Nations Sustainable Development*. (2015).https://www.un.org/sustainabledevelopment/sustainable-development-goals/, accessed on 1st September, 2022.

Vetro, A., Canova, L., Torchiano, M., Minotas, C.O., Lemma, R., Morando, F. Open data quality measurement framework: Definition and application to open government data. Government Information Quarterly. **33**(2), 325-337. (2016). https://doi.org/10.1016/j.giq.2016.02.001.

Wiemann, S., Bernard, L. Spatial data fusion in spatial data infrastructures using linked data. International Journal of Geographical Science. **30**(4), 613-636. (2016). https://doi.org/10.1080/13658816.2015.1084420.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. **3**, 160018. (2016). https://doi.org/10.1038/sdata.2016.18.

Williams, A.J., Harland, L., Groth, P. et al. Open PHACTS: Semantic interoperability for drug discovery. Drug Discovery Today. **17**(21-22), 1188-1198. (2012). https://doi.org/10.1016/j.drudis.2012.05.016.

World Bank GCC economies expected to expand by 6.9% in 2022. (2023). Available at https://www.worldbank.org/en/news/press-release/2022/10/31/gcc-economies-expected-to-expand-by-6-9-in-2022, accessed on 16th January, 2023.

Zhou, K., Ethayarajh, K., Card, D., Jurafsky, D. Problems with cosine as a measure of

embedding similarity for high frequency words. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. **2**, 401-423. (2022). https://aclanthology.org/2022.acl-short.45.pdf.

Zuiderwijk, A., Janssen, M., Davis, C. Innovation with open data: Essential elements of open data ecosystems. Information Polity. **19**(1-2), 17-33. (2014). https://doi.org/10.3233/IP-140329.

Gurin, J., Manley, L., Ariss A., Sustainable Development Goals and Open Data, World Bank. (2015). Available at https://blogs.worldbank.org/en/digital-development/sustainable-development-goals-and-open-data#:~:text=Open%20Data%20can%20help%20achieve,for%20action%20on%20national%20issues. Accessed on 22nd July 2024

Petrov, O.V., Gurin, J., & Manley, L. Open Data for Sustainable Development. (2016). https://www.semanticscholar.org/paper/Open-Data-for-Sustainable-Development-Petrov-Gurin/7d6582b5cf940f0f0e87333cf2b4047230109513

Palacios, J.E., & Piedra, N. Connecting Open Data and Sustainable Development Goals using a Semantic Knowledge Graph Approach. *ONTOBRAS*. (2019).

Jaiswal, A., A study on role of open government data towards achieving sustainable development goals in india. Journal of Harmonized Research in Management. 5. 29. (2019). 10.30876/JOHR.5.2.2019.29-32.

Chokki, A.P., Alexopoulos, C., Saxena, S., Frénay, B., Vanderose, B. and Ali, M., "Metadata quality matters in open government data (OGD) evaluation! An empirical investigation of OGD portals of the GCC constituents", Transforming Government: People, Process and Policy, Vol. 17 No. 3, pp. 303-316. (2023). https://doi.org/10.1108/TG-09-2022-0118

Alromaih, N., Albassam, H., and Al-Khalifa, H. A proposed checklist for the technical maturity of open government data: an application on GCC countries. In Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services (iiWAS '16). Association for Computing Machinery, New York, NY, USA, 494–499. (2016). https://doi.org/10.1145/3011141.3011211

Ali, M., Alexopoulos, C., and Charalabidis, Y. A comprehensive review of open data platforms, prevalent technologies, and functionalities. In Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance (ICEGOV '22). (2022). Association for Computing Machinery, New York, NY, USA, 203–214. https://doi.org/10.1145/3560107.3560142