

Received March 16, 2022, accepted May 6, 2022, date of publication May 10, 2022, date of current version May 16, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3174090

Toward Building a Linked Open Data Cloud to Predict and Regulate Social Relations in the Saudi Society

AFNAN MUHAMMADSALH ALSUKHAYRI¹, MUHAMMAD AHTISHAM ASLAM¹,
IMTIAZ HUSSAIN KHAN², RABEEH AYAZ ABBASI³, AND AMAL BABOUR¹

¹Department of Information System, King Abdulaziz University, Jeddah 22254, Saudi Arabia

²School of Technologies, Cardiff Metropolitan University, Cardiff CF5 2YB, U.K.

³Department of Computer Science, Quaid-i-Azam University, Islamabad 45320, Pakistan

Corresponding author: Muhammad Ahtisham Aslam (maaslam@kau.edu.sa)

ABSTRACT

Background: The trend in producing linked open data to publish high-quality interlinked data has gained widespread traction in recent years. Various sectors are producing linked open data to increase public access and ensure transparency, in addition to a better utilization of government data, namely linked open government data.

Problem Definition: As compared to the developed countries, Saudi Arabia lags behind in benefiting from this new era of ubiquitous web of data, despite its publication of government related data in non-linked format. In the context of Saudi open government data, the full potential of multi-category data published by various government agencies at different portals is not being realized as the data are not published in open data format and remain unlinked to other existing datasets.

Methodology: To bridge this gap, this study presents a framework to extract and generate semantically enriched data from various data sources under different domains. The framework was used to produce the Saudi linked open government data cloud by interlinking data entities with each other and with external existing open datasets.

Results: The effectiveness of our approach is validated by applying it to a socially significant issue, i.e., divorce rate, in Saudi Arabia. By posing smart queries to semantically enriched data, we were able to perform an in-depth analysis of different factors related to increasing divorce rates in Saudi Arabia. Arguably, without using linked open data and related technologies such analysis would not have been possible. Finally, we also present a simulated visual environment for better understanding and communication of such analysis for decision and policy makers.

INDEX TERMS Linked open government data, smart queries, machine reasoning, ontologies, semantic technologies.

I. INTRODUCTION

The linked open data (LOD) concept is used to retrieve structured data from different and distributed data sources. Linked open data is important in information and data management field because it is usually preferred over traditional data management. Originating in Europe, Australia, New Zealand, and North America, the open government data movement has begun to gain momentum in Asia, South America, and Africa. Currently, the countries which are leading open government data activities include the US, UK, Australia, and

the Scandinavian countries [10]. They have an open data community and commitment to both open data and central open data portals. However, in the Saudi context, open government data are published in silence as it is neither made available in an open data format nor linked to other existing datasets. Thus, there is an urgent need to produce and publish these data in a more structured and machine understandable format by using Resource Description Framework (RDF), to support more nuanced queries by making use of SPARQL Protocol and RDF Query Language (SPARQL). Open data is an important aspect of transparent governance, as it includes not only government data but also data from various domains such as business, industry, citizen, science and education.

The associate editor coordinating the review of this manuscript and approving it for publication was Claudio Zunino.

The world leading institutions which broadly use open data are the World Bank, United Nations, New York Times and The Guardian and Open Knowledge Foundation [10].

Typically, governments lead the initiative to produce open government data. The Saudi Arabian government has already taken an initiative towards producing open data but generating and utilizing linked open government data are still challenging areas for the Saudi government to address. The Saudi Arabian government developed the open data portal¹ which is aimed at enabling transparency, promoting e-participation and making informed decisions about government policies. The open data portal also allows the public to access, search, download, and use datasets published by different governmental departments and ministries. However, the data are currently available in various formats including Excel sheets, databases, documents, and Comma Separated Values (CSV) files, located in different data sources and are of poor quality, which makes it difficult for these to be used for effective analysis by policymakers to fulfil the required values. So, generating policy insights would be difficult without establishing linked open data and providing SPARQL endpoints.

In addition, during the last 40 years, Saudi Arabia has experienced rapid social transformation, economic development, education, labour forces, and changes in the roles and status of women [5], [15]. This situation calls for a transformation of open government data into linked open government data. This is likely to benefit organizations, governments, and individuals by improving transparency, encouraging public participation and creating the ability to query multiple data sources [6], [21]. A number of studies such as [10] and [9] show that linking open government data facilitates transparency, citizen awareness, accountability, in addition to supporting better decision-making, avoiding duplication and enabling cost saving in data collection. Other research such as [9] and [3] also suggests that linking open government data from different resources can improve the economy, offer social value, strengthen democracy and reduce the cost of public functions. We are interested in creating and using Saudi linked open data cloud to enable the posing of more detailed and comprehensive queries for informed decision making.

To the best of our knowledge, limited research has been conducted related to producing and making the computational use of linked open data in Saudi context [3]. This is the first study of its kind that created linked open data as well as used it to uncover various factors that have impact on social relations in Saudi society. Building on our previous work [3], [4], this paper presents an extended framework to publish Saudi linked open government data. We identified and used different data sources from various domains to produce Saudi linked open data which contains interlinked data about population, household income, educational status, job opportunities, household expenditure, consumption and living costs. The datasets for this research are primarily from the General Authority for Statistics, which publishes publicly

accessible datasets from governmental departments and ministries. We used the Ministry of Education, DBpedia, World Bank, and the Ministry of Justice as data sources that can be used to link various open datasets published in the linked open data cloud. The ultimate goal is to use this interlinked data to investigate different factors and parameters that have direct or indirect impact on social aspects of Saudi society and then to use results of such studies to define policies to address social issues in an effective way.

As a case study, we applied our linked open data approach to identify the underlying reasons in the increased divorce rates in Saudi Arabia. Since the 1990s, the divorce rates in Saudi Arabia have risen from 25 percent to 60 percent [5]. The present research provides new insights into understanding the reasons for increase in divorce rates in Saudi Arabia. We explored the potential impact of age structure, household income/expenditure, educational status, job opportunities, and other factors on divorce rates in Saudi Arabia. The results can help decision makers in Saudi Arabia to identify effective solutions for the underlying factors and parameters to decrease divorce rates in the country. We also establish a SPARQL endpoint where users can pose various queries to Saudi linked open data in order to retrieve and analyze the results. This research extends and enriches our previous work [3], [4] in the following way:

- Explored and identified different data sources from various domains that have direct or indirect impact on the social relations in Saudi society.
- Enhanced framework names as Saudi Linked Open Data Cloud Framework (SLODCF) that is used to identify, extract, process and produce semantically enriched data (in resource description framework, henceforth, RDF, format) from various data sources.
- Enriched Saudi linked open government data cloud consisting of interlinked RDF datasets which are linked to each other as well as with external existing open datasets such as Ministry of Education, Ministry of Justice, General Authority for Statistics, DBpedia, and World Bank.
- Extended SPARQL endpoint to pose semantically enriched queries by making use of SPARQL protocol.
- A case study of using the Saudi linked open government data cloud for computational interpretation of underlying factors that have direct or indirect impact on social relations in Saudi society.
- Visualizing the Saudi linked open government data using a dashboard to better understand and communicate with large amounts of data.

The rest of the paper is organized as follows. Section II discusses the related work. The process of identifying various data sources and the underlying factors and parameters that impact social relations is discussed in Section III. The framework that we developed and used to identify, extract, process and produce semantically enriched data (in RDF format) from various data sources is outlined in Section IV. In Section V, we present the Saudi linked open data cloud followed by

¹<http://www.data.gov.sa/>

SPARQL endpoint in Section VI. Section VII presents the case study to demonstrate the effectiveness of our approach. In Section VIII, we present simulated visual environment to query, visualize, and analyze data from Saudi linked open government data cloud for policy and decision making which is discussed in Section IX. The paper concludes in Section X.

II. RELATED WORK

A plethora of work has been published on linked open data in various domains such as governance, music and medicine. In recent years, a number of governments have led the initiative to make their data freely available on the Web for everyone to access, reuse, redistribute and share for their purposes [2], [9], [28]. This is beneficial to organizations, governments, and individuals [9]. In [26], the authors have demonstrated that linking open government data from different resources can improve the economy, create social value, strengthen democracy and reduce the cost of public functions.

The authors in [17] developed a web application to retrieve and consume linked open data. The web application prototype solved the problem of open data distribution by combining and visualizing these data using linked open data technologies. They used data sources related to education and job vacancies in Jakarta. This web application enables data storage, auto update of data, query, web scraping and dashboard visualization. Major constraint of this application are as follows: First, updated obtained data from open data portal are not provided. Second, this application involves only two web sites as data sources (this made the scope of work very limited). Third, this application provides very limited dashboard which is not be enough for data querying as well as statistical analysis of data.

In [20], the authors proposed a framework to link and query open datasets from various government portals to generate graph based on RDF data. The authors made use of an online portal as source for open data and made the extracted datasets available as open data. Main limitations of this work are (a) it did not interlink the datasets (i.e., no linked open data provided), (b) the available SPARQL endpoint can be used to query a specific datasets which constrains the utilization and analysis of data as linked data.

A framework named as Framework of Thai Local Government System has been presented in [11]. This framework is developed to maximize the interaction and data integration between different department at Sub-district Administrative Organization of Thai government. It allowed accessing, discovering, and achieving efficient interoperability of information systems. Additionally, this system also facilitates to query the RDF data by making use of the SPARQL protocol. In this work, authors used the road construction project as a case study for their project. Also, in this work authors confined their work by depending on existing tools such as D2R and Drupal in extracting, producing, and publishing the open data.

In the medical domain, some studies have proposed diverse approaches to linked datasets from medical sources.

In a number of studies [22], [24], the authors solved the problem of distributed data on the Web by developing a methodology to generate linked data in drug product domain. In [22], the authors used datasets from global drug product data, transforming and publishing consolidated drug product data as linked drug data. In [24], the authors integrated open data sources relevant to clinical trials such as LinkedCT clinical trials dataset, the SIDER drugs and side effects dataset using linked data technology and developed an intelligent patient-centred clinical trial decision-support system. This provides automated reasoning in search, analysis and decision making. However, with all these benefits, these two approaches have some major limitations such as (a) the researchers in this study did not conduct a user evaluation for their prototype, (b) the datasets that are included in the system are very limited resulting into incomplete reasoning and results, (c) very limited identification of the factors and parameters that can be useful for clinical trials of different patients.

In [1], a novel hybrid approach is proposed which employs named entity recognition/disambiguation and dependency parsing integrated with Wikidata ontology to develop an Arabic question-answering system. They achieved an above 80% accuracy. In a recent study [27], the authors discussed ca. 80 papers on the assessment of linked open data. Even though, authors claim that they have developed the Arabic QA System but they acknowledged the limitation of their work as having very limited dataset. At the same time authors acknowledged the limited support of Arabic language in the global LOD. Also, this system offers very limited queries such that it supports up to only the first level of properties as well as with limited support of aggregation functions such as "ORDER BY".

Similarly, a "group recommender system" is presented in [25]. The proposed system uses a linked open data model to ameliorate data sparsity problem. Also, this work is a joint venture of implementing aggregation strategies, group size as well as semantic technologies to maximize the accuracy and precision of the recommendations given by the proposed system. Similarly, in [31], authors have provided a systematic literature review on linked open data in location-based recommendation systems for tourism. Even though the focus of [31] is tourism but its datasets cause the limitations in the recommendations. For example, major limitations of their work include ignorance of many key factors/features such as trust, friendship, and context of information. Due to unavailability of these factors in the recommendations, the proposed system loses the trust level of individuals as well as groups.

Our work addressed most of the limitations of existing works that we highlighted above. The major limitations of the existing work that we addressed in this study are: (a) using multiple local as well as international data source, (b) our own developed framework to extract and produce semantically enriched datasets from multiple sources in various domains (c) our framework can be customized to process incremental data sources (d) evaluation, results and analysis of the

proposed system with a real life case study (e) simulated environment with updated datasets integrated with SPARQL endpoint, and (f) interactive dashboard with visual effects making it easy for domain experts and stakeholders to visualize and analyze data easily.

III. IDENTIFICATION OF DATA SOURCES

In this new era of “Web of Data” [16], we can find lot of sources of data and information over the Internet. Searching and identifying right resources of data can be very helpful in achieving a goal that is very much data dependent and data oriented. Keeping this principle in front, we explored lot of data sources over the Internet and identified data sources from various domains such as the Ministry of Education, Ministry of Interior, Ministry of Justice, General Authority for Statistics, DBpedia, World Bank, and Human Resource and Social Development to extract the data from and to make it part of our Saudi linked open data cloud. The identification and classification of above-mentioned sources is based on the data that they have related to different social aspect of society which are also related to the presented case study (i.e., analyzing causes of divorce rate in Saudi society). Also, data sets are classified as reliable such as World Bank, General Authority for Statistics as well as public datasets such as Twitter data. In what follows, we describe some of these data sources, their key factors and parameters, especially that are related to our case study (i.e. social relations in Saudi society).

A. GENERAL AUTHORITY FOR STATISTICS

The General Authority for Statistics² is a government agency in Saudi Arabia responsible for the implementation of statistical work including the national surveys. It is an innovative statistical reference for Saudi Arabia’s socio-economic development. The datasets exist in Excel sheets, databases, and CSV formats. So, in our work, we used datasets related to populations, education, incomes, and job statistics. These datasets are available in different formats and describe the population based on five dimensions: 1) Age Groups, 2) Gender: male and female, 3) Employee Income, 4) Unemployment Rates, and 5) Administrative Areas in Saudi Arabia. The administrative areas are: Al-Riyadh, Makkah Al-Mokarramah, Al-Madinah Al-Monawarah, Al-Qaseem, Eastern Region, Aseer, Tabouk, Hail, Northern Borders, Jazan, Najran, Al-Baha, and Al-Jouf.

B. MINISTRY OF EDUCATION

The Ministry of Education³ aims at an exceptional education system that builds a globally competitive knowledge-based community. These datasets are available in different formats. The collected datasets describe population based on four dimensions: 1) Gender: male and female, 2) nationality: Saudi and non-Saudi, 3) Marital Status: Married, Divorced, Widowed and Never Married, 4) The same administrative

areas in Saudi Arabia as for General Authority of Statistics, and 5) Educational Status: primary, intermediate, secondary, university, master and Ph.D.

C. DBPEDIA

DBpedia includes structured information from Wikipedia that is available on the Web [8]. DBpedia provides real data from various domains [8]. It allows queries generation for Wikipedia and linking to other datasets on the Web [8]. The Saudi DBpedia datasets are available in an open data format. We linked our Saudi open datasets (such as population) with the Saudi DBpedia dataset. For example, we used datasets about Makkah Al-Mokarramah region and other regions in Saudi Arabia.

D. WORLD BANK

World bank data⁴ include global development data for different indicators pertaining to all world countries. The Saudi World Bank dataset includes more than 1500 indicators. In our research, we used the Saudi World Bank dataset. For example, we included indicators pertaining to the percentage of annual household consumption expenditure per capita, technology usage, the average monthly household expenditure for divorced and married people and the general living costs per year.

E. MINISTRY OF JUSTICE

The role of the Ministry of Justice⁵ is to monitor the Saudi courts and to ensure that financial and administrative requirements are fulfilled. We used datasets on general indicators of requests for certification of documents such as marriage requests and divorce requests.

We collected the data from the afore-mentioned sources. Table 1 summarizes the category of datasets, the number of datasets, and the source of the datasets. These data sources are used as an input for the SLODCF framework to process and produce semantically enriched data (as discussed in the next section).

IV. SAUDI LINKED OPEN DATA CLOUD FRAMEWORK (SLODCF)

This section describes the enhanced SLODCF framework. Building on our previous work [3], [4], we enhanced Saudi linked open government data cloud framework (SLOGDCF) to Saudi linked open data cloud framework (SLODCF). The ultimate update is that the current framework efficiently processes and manipulates not only the government data but also the data produced by third parties such World Bank and social media sites (e.g., Twitter). The overall process of SLODCF framework is that it collects, processes, generates RDF datasets, interlinks these datasets with each other and with other open datasets (published from government and non-government resources). The SLODCF framework consists of

²<https://www.stats.gov.sa/en>

³<https://www.moe.gov.sa/en/default.aspx>

⁴<https://www.worldbank.org/>

⁵<https://www.moj.gov.sa/english/pages/default.aspx>

TABLE 1. Saudi linked open data cloud datasets statistics.

Category	Number of datasets	of Source datasets	of
Population, Education	136	Ministry of Education	of
Population, Incomes and Job Statistics	236	General Authority for Statistics	for
Certification of Documents	7	Ministry of Justice	of
Population, household consumption expenditure, the general living costs Saudi Arabia Administrative Areas	247	World Bank DBpedia	

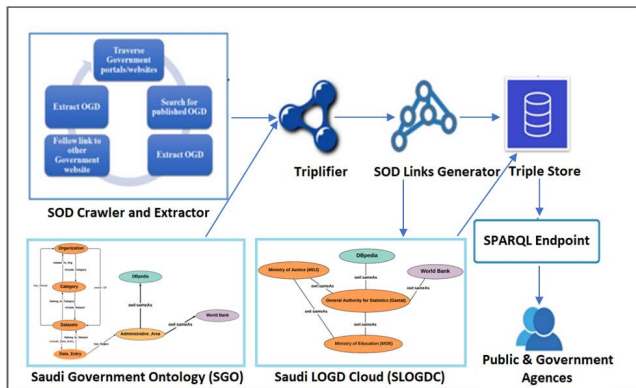


FIGURE 1. Saudi linked open government data cloud framework (SLOGDCF).

four modules: data preparation, data modelling, data linking and querying (as shown in Fig. 1). Here, we describe these modules in detail.

- Saudi Open Data (SOD) Crawler and Extractor:** This is the main module of our framework. It traverses, collects, and extracts Saudi open data from multiple data sources such as the Ministry of Education, DBpedia, World Bank, General Authority for Statistics, and the Ministry of Justice. Our framework traverses open data available on Saudi ministries’ websites and portals and extracts the required data entities. It follows the links between data objects in a way that the connected data are also retrieved. Linking open datasets requires investigation between government organizations supervised by certain ministries. This process of finding and extracting datasets is adapted from the indigenous updated algorithm (i.e., Algorithm 1). The Algorithm 1 is an extension of our initial work (as discussed in [3]) for crawling and extracting Saudi open data. It can be

Algorithm 1 Data Identification, Processing and Triplifying Algorithm.

```

Data: Source portals (government and non-government bodies) and social media data sources such as Twitter
Result: SPARQL Endpoint with resulting RDF datasets in N-Triple format at the back end
1   $N - Triples = \{ \} ;$ 
2   $S = \{ \text{Source Website/portals, CSV files \& social media sites } S \} ;$ 
3  foreach government and non-government data source  $s \in S$  do
4       $F = \{ \text{Categorization and identification of data sources containing social factors and parameters } s \} ;$ 
5      foreach category  $f \in F$  do
6           $F = \{ \text{Set of various factors in category } f \} ;$ 
7          foreach data source  $d \in D$  do
8               $K = \{ \text{Set of all parent factors } f \} ;$ 
9              foreach data source  $k \in K$  do
10                  $Triples = ParentsourceN - Triples(triplification) \cup ChildsourceN - Triples(k);$ 
11                  $L = \{ \text{links to the child data sources } k \} ;$ 
12                 foreach child data sources  $l \in L$  do
13                      $Triples = ParentsourceN - Triples(triplification) \cup ChildsourceN - Triples(k);$ 
14                 end
15             end
16         end
17     end
18     Generate RDF datasets by using Triples;
19 end

```

used to find, extract, link and publish the interlinked datasets from different afore-mentioned (i.e., government, non-government, and social media resources) data sources. The Algorithm 1 is improved with capacity of identifying the underlying factors together with their related data to be extracted from identified resources. An improved feature of this algorithm is that it helps the SLOGDCF framework to starting crawling and processing from parent source and move up to last nested resource by using recursive algorithm. This recursive approach increases the data processing and collection capacity of the SLOGDCF framework which ultimately resulted into bigger knowledge graph support better and deeper analysis of underlying factors.

After extracting datasets from different data sources, we modelled our data to create an RDF model and to generate RDF datasets. Here, we show some examples of RDF datasets from Saudi open government data by using SLOGDCF. This will result in RDF triples which are in turn used to produce RDF datasets in

TABLE 2. Description of some properties extracted from Saudi open government data.

Property	Description	Example
age_Group	Provide data about the age group for a specific data entry.	AgeGroups_MaritalStatus age_Group "60-64"
educationalStatus	Provide data about the education status for a specific data entry.	Education_MaritalStatus educationalStatus "Master"
year	Provide data about the year for a specific data entry.	AgeGroups_MaritalStatus year 2017
has_Region	Provide data about the region for a specific data entry.	EducationalStatus_Region has_Region Aseer

N-triple formats. Therefore, we extracted information on Saudi datasets from DBpedia datasets including different attributes such as total population and more related information regarding Saudi regions. To extract structured information from these data sources which are available in different data formats, we used different generation tools and techniques. So, the results of structured information extraction pertain to datasets about population, household income, educational status, job opportunities, household expenditure, consumption, and living costs from different data sources. Table 2 shows an example of some properties extracted from Saudi open government data and Table 3 shows some examples of RDF statements belonging to Saudi open government data datasets.

- **Triplifier:** The triplifier module is used to process and produce RDF triples in the form of Subject, Predicate, and, Object. Before this process, the triplifier also checks the extracted data entities, their types and automatically classify them as classes or properties of the Saudi Government Ontology (SGO). After classification the extracted data entities are mapped as instances of specific classes or as properties which are used to link instances of two classes or link instances of classes with literal values. These classified instance and properties become the part of final datasets. This module also uses external tools and APIs such as OWL APIs to generate RDF datasets in N-triple format.
- **Saudi Open Data (SOD) Links Generator:** This module uses the RDF datasets to link the open datasets. First, it determines the type of data and their relations by

TABLE 3. Some examples of RDF statements generated from Saudi open government data.

Subject	Predicate	Object
AgeGroups_MaritalStatus	age_Group	"60-64"
Education_MaritalStatus	educationalStatus	"Master"
AgeGroups_MaritalStatus	year	2017
EducationalStatus_Region	has_Region	Aseer
Household_Income	married_income	13017

examining the information that describes datasets. Then, established more relations between data entities within datasets. The data items extracted from data sources contain data related to the presented case-study. The creation and management of relation types between data entities within datasets or links between data sources are identified and created semi-automatically. Then the RDF links are automatically generated between these data entities to one another to create Saudi linked open data, henceforth creating bigger knowledge graphs.

The internal and external links improve usability by connecting the Saudi open government data and other data sources with the LOD cloud. External links enables easier browsing of the linked open data cloud and Saudi linked open government data by following the links. The internal links established between Saudi open government data to create Saudi linked open government data, the links created if these datasets related to each other, or if these datasets belong to the same organization/category or belong to a different organization/category. The external links established to create Saudi linked open government data cloud, external links created between Saudi linked open government data and other existing open datasets. Table 4 lists the datasets and the number of links between these datasets.

- **Saudi Linked Open Data SPARQL Endpoint:** This module acts as a front end for end users. It enables researchers to put semantically enriched SPARQL queries to Saudi linked open datasets through the SPARQL endpoint. Saudi linked open data SPARQL endpoint was used to run various queries and to analyze the obtained results while addressing various factors according to our case study.

V. SAUDI LINKED OPEN GOVERNMENT DATA CLOUD

In order to enable Saudi linked open data users to retrieve details and further information, we interlinked Saudi open datasets with one another and with other existing open datasets to create the Saudi linked open government data cloud. Once the data model of Saudi open government data is created and RDF datasets are generated, we linked the data entities within our datasets and other existing open

TABLE 4. Number of links between datasets in Saudi linked open government data cloud (SLOGDC).

Dataset	Link Count	Dataset
GaStat	28,230	World Bank
MOE	19,300	GaStat
MOI	18,520	GaStat
GaStat, MOJ, MOI	14,450	Twitter
MOJ	14,390	GaStat
GaStat	13,230	DBpedia
GaStat	12,230	HRSD, MCS
MOJ	8,460	MOE
MOI	8,450	MOJ
Total	137,260	

GaStat: General Authority for Statistics
 MOJ: Ministry of Justice
 MOE: Ministry of Education
 MOI: Ministry of Interior

datasets as discussed above. In order to do that, we identified the entities related to Saudi Arabia and related to our data and then identified the relationships between entities in our datasets and other datasets from different data sources using interlink tools and OWL properties. This linking approach allows access and retrieval of detailed information, resulting in larger knowledge graph from different connected datasets belonging to different domains.

A. GENERATING SAUDI GOVERNMENT ONTOLOGY (SGO) FOR SAUDI LINKED OPEN GOVERNMENT DATA CLOUD

In order to convert our datasets from different formats to RDF/OWL, we developed our own ontology (and reused existing ontologies wherever possible as a rule of best practices in ontology development) and RDF triples from different formats by using ontology development tools [23], [30]. Our ontology includes several classes such as *Organization*, *Category*, *Datasets*, and *Data Entry* as shown in Fig. 2. The *Datasets* class represents the statistical datasets which are the container of data such as population datasets and economic datasets. The *Data Entry* class represents a single piece of data that represents the population by gender and marital status identifiable by a set of dimensions such as region, age group, and years. The *Organization* class represents the government agencies or private agencies that published open datasets. The *Category* class represents the category or type of the dataset.

In addition, our ontology includes several properties as shown in Fig. 2. The *include_Category* property relates the organization to the category. The *owner_Of* property relates the organization to the dataset. The *belong_to_Category* property relates the dataset to the category. The *has_Owner* property relates the dataset to the organization. The *related_To_Org* property relates the category to the organization. The *include_Dataset* property relates the category to the dataset. The *belong_to_Dataset* property relates the data entry to the dataset. Some of the classes such as Data Entry may contain subclasses depending on the case studies and

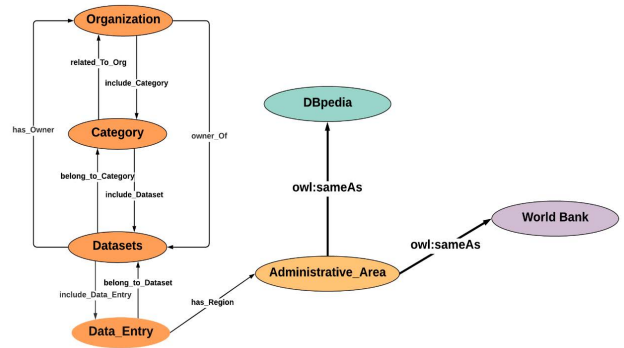


FIGURE 2. An overview of Saudi government ontology.

datasets which are generated automatically when we generate our ontology and RDF triples from different formats.

B. LINKING DATASETS TO CREATE SAUDI LINKED OPEN GOVERNMENT DATA CLOUD

The linked open data is a set of the best practices that can be used for publishing, interlinking and querying data from distributed and different data sources [7]. The dataset of the population of people in General Authority for Statistics includes common attributes related to the population such as marital status, gender, region, and nationality. However, they do not include other attributes in different domains such as educational status, job opportunities, age groups, incomes, household expenditure, consumption and living costs. To enrich our current datasets, we collected data from external data sources such as Ministry of Education, DBpedia, World Bank, and Ministry of Justice.

RDF that is used to represent open data as linked open data consists of nodes and directed edges for connecting the nodes. A Uniform Resource Identifier (URI) uniquely identifies an individual node. RDF triples are interlinked by using the object as the subject for triple or by using a new object for existing subjects as shown in Fig. 3. The aggregation of many interlinked RDF triples contributes to creating an RDF graph. Linked open data principles foster the use of linked open government data because it makes the aggregation and interlinking of heterogeneous data from different sources much easier. Using linked open data principles, there is no need for linked open data consumers to learn different data access techniques for different data sources.

Therefore, we used a linked data approach to link one data source to another, which allows users to explore more information from different data sources and to get proper data which are otherwise not available through isolated datasets [4], [12]. So, we created a Saudi linked open government data cloud that can be used to provide new insights into different problems including divorce rates issues in Saudi Arabia and propose effective solutions. Before we link to other data sources, we first need to choose the target external data sources that can enhance our existing datasets and then link these to the target data sources. We also provide SPARQL endpoint where users can pose different types of queries to

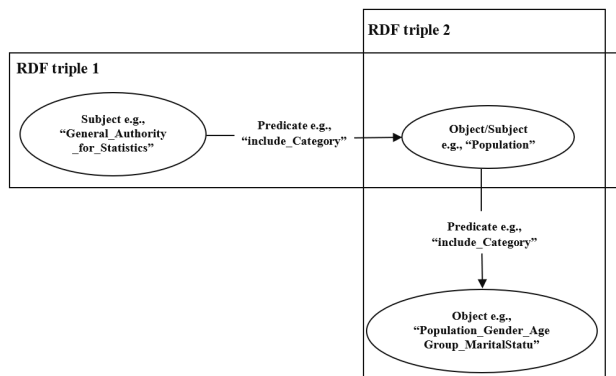


FIGURE 3. Two interlinked RDF triples.

retrieve the results and analyze them. To link our datasets to other open datasets in the Saudi linked open government data cloud, we used interlinking approaches to link entities from our datasets to other entities from other datasets as shown in Fig. 4. The resulting datasets include various interlinked data sources belonging to diverse domains and governmental departments.

This study analyzes the RDF datasets for determining links under certain conditions. The Saudi open government data interlinking functionality of linked open data is essential to create Saudi linked open government data cloud. Linking of Saudi open government data is done by the RDF mechanism with the option to choose from two types of data linking: internal and external links. Fig. 4 represents the types of data linking at an abstract level.

The internal and external links improve the usability by connecting the Saudi open government data and other data sources with the linked open data cloud. External links enhance connectivity among different data sources. The internal links are established between Saudi open government data to create Saudi linked open government data, the links created if these datasets related to each other, or if these datasets belong to the same organization/category or belong to a different organization/category. The external links are established between Saudi linked open government data and other existing open datasets to create Saudi linked open government data cloud. Therefore, we were able to retrieve maximum data and to produce bigger knowledge graphs about the population and demographic characteristics in Saudi Arabia not only from our indigenous dataset but also from all the other datasets which were interlinked to it.

VI. SAUDI LINKED OPEN GOVERNMENT DATA SPARQL ENDPOINT

The SPARQL endpoint can be used to pose queries by making use of SPARQL protocol which is not possible by using traditional data. The SPARQL endpoint is a flexible way to access and query linked open data [7]. Here, we provide a SPARQL endpoint for querying the Saudi linked open government data cloud which includes different data sources such as General Authority for Statistics, Ministry

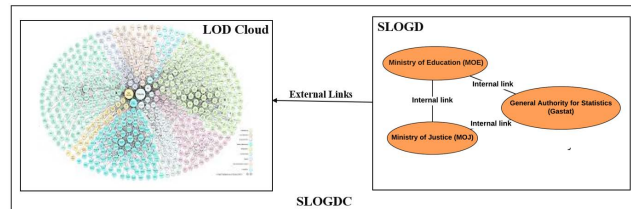


FIGURE 4. Saudi linked open government data cloud.

of Education, DBpedia, World Bank, and Ministry of Justice. We make use of Saudi linked open government data cloud to query various factors under multiple domains, such, as the Saudi population, educational status, job opportunities, age groups, incomes, household expenditure, consumption, and living costs to identify divorce rates in Saudi Arabia. In addition, we present some samples of SPARQL queries executed through SPARQL endpoint to assess the quality of Saudi linked open government data cloud as shown in Figures 5–8. These queries integrated data from multiple domains. These queries retrieved data from our RDF datasets and then obtained data from multiple open datasets in the Saudi linked open government data cloud in order to get more information as well as related information, which would not be possible if we had not created the Saudi linked open government data cloud. The queries shown in Fig. 5 to Fig. 8 relate the total number of divorced males/females according to a variety of factors in Saudi Arabia per year (i.e., 2017) to find divorce rates. The query shown in Fig. 5 indicates the educational level of Saudi divorced males and females to find the education of most divorced people by retrieving the parameters of educational status and the number of divorced males and females in every educational status. In Fig. 6, the query indicates Saudi population by region for divorced males and females so as to find the region that has the highest number of divorced people and it retrieves the parameter region and the number of divorced males and females in each region. In Fig. 7, the query indicates the total unemployment rates of Saudi males and females by their educational level and regions so as to find the unemployment rate for divorced people based on their education and region and the region with the greatest number of divorced population members retrieved from the previous queries in Fig. 5 and Fig. 6. Within the set parameters, it retrieves the total unemployment rates for males and females based on their education and region. In Fig. 8, the query indicates the average monthly wages per paid employee for Saudi males and females to find the income for divorced people based on their education retrieved from the previous query in Fig. 5. In the context of the parameters, it retrieves the income for males and females based on their education.

Similarly, mutatis mutandis, we developed queries to compute the Saudi population for divorced males and females by their age groups, the annual percentage growth of household and final consumption expenditure per capita, and the the population which is covered by health insurance.


```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX GaStat: <http://www.semanticweb.org/afnan/ontologies/untitled-ontology-62#>
PREFIX SGO: <http://www.semanticweb.org/afnan/ontologies/untitled-ontology-59#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX MOI: <http://www.semanticweb.org/afnan/ontologies/2020/8/untitled-ontology-60#>
PREFIX MOJ: <http://www.semanticweb.org/afnan/ontologies/untitled-ontology-64#>

SELECT ?Educational_Status ?Number_of_Divorced_Male ?Number_of_Divorced_Female
WHERE {
?Data_Entry rdf:type GaStat:Education_Marital_Status;
GaStat:educational_Status ?Educational_Status;
GaStat:year 2017;
GaStat:divorced_Males ?Number_of_Divorced_Male ;
GaStat:divorced_Females ?Number_of_Divorced_Female.}

```

FIGURE 5. Sample query for educational status of divorce people.

```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX GaStat: <http://www.semanticweb.org/afnan/ontologies/untitled-ontology-62#>
PREFIX SGO: <http://www.semanticweb.org/afnan/ontologies/untitled-ontology-59#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX MOI: <http://www.semanticweb.org/afnan/ontologies/2020/8/untitled-ontology-60#>
PREFIX MOJ: <http://www.semanticweb.org/afnan/ontologies/untitled-ontology-64#>

SELECT ?Region ?Number_of_Divorced_Male ?Number_of_Divorced_Female
WHERE {
?Data_Entry rdf:type GaStat:Region_Marital_Status;
GaStat:has_Region ?Region;
GaStat:year 2017;
GaStat:divorced_Males ?Number_of_Divorced_Male ;
GaStat:divorced_Females ?Number_of_Divorced_Female.}

```

FIGURE 6. Sample query for divorce people by region.

The above queries enabled us to retrieve useful information from Saudi linked open government data cloud that we created as part of this study. Detailed analysis and discussion of these results follows in the next section.

VII. COMPUTATIONAL INTERPRETATION OF DIVORCE RATE FROM SAUDI LINKED OPEN GOVERNMENT DATA CLOUD

In this section, we present the quantitative and qualitative analysis of the resulted data. There are many demographics, social and economic variables which influence the divorce rates in Saudi Arabia [13]. After querying Saudi linked open government data cloud and retrieving the results, we were able to conduct a more detailed and comprehensive analysis to obtain results on divorce rates in Saudi Arabia which can be used to define future policies and decision making.

A. SAUDI REGIONS-BASED ANALYSIS

The data revealed that the Makkah Al-Mokarramah region has the highest divorced male and female population amongst other regions where the divorced female population numbers ca. 50,000 and divorced male population numbers ca. 20,000. This is followed by the Riyadh region which also has a high divorced male/female population amongst the other regions, with a divorced female population ca. 30,000 and divorced male population ca. 20,000. Also, the region with the lowest divorce rates among other regions is the Northern border region.

Therefore, to do this analysis properly we use the Crude Divorce Rate (CDR) to capture the changes in divorce rates because the data needed for measuring the CDR are retrieved by querying the Saudi linked open data cloud through SPARQL endpoint. The CDR is the number of divorces in

```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX GaStat: <http://www.semanticweb.org/afnan/ontologies/untitled-ontology-62#>
PREFIX SGO: <http://www.semanticweb.org/afnan/ontologies/untitled-ontology-59#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX MOI: <http://www.semanticweb.org/afnan/ontologies/2020/8/untitled-ontology-60#>
PREFIX MOJ: <http://www.semanticweb.org/afnan/ontologies/untitled-ontology-64#>

SELECT ?Region ?Unemployment_Rate_Male ?Unemployment_Rate_Female
WHERE {
?Data_Entry rdf:type GaStat:Unemployment_Rate_Region;
GaStat:has_Region ?Region;
GaStat:year 2017;
GaStat:total_Males ?Unemployment_Rate_Male ;
GaStat:total_Females ?Unemployment_Rate_Female.}

```

FIGURE 7. Sample query for unemployment rates and average wages of divorce people by educational status.

```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX GaStat: <http://www.semanticweb.org/afnan/ontologies/untitled-ontology-62#>
PREFIX SGO: <http://www.semanticweb.org/afnan/ontologies/untitled-ontology-59#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX MOI: <http://www.semanticweb.org/afnan/ontologies/2020/8/untitled-ontology-60#>
PREFIX MOJ: <http://www.semanticweb.org/afnan/ontologies/untitled-ontology-64#>

SELECT ?Educational_Status ?Income_of_Male ?Income_of_Female
WHERE {
?Data_Entry rdf:type GaStat:Income_Educational_Status;
GaStat:educational_Status ?Educational_Status;
GaStat:year 2017;
GaStat:total_Males ?Income_of_Male ;
GaStat:total_Females ?Income_of_Female.}

```

FIGURE 8. Sample query for average wages of divorce people by educational status.

a given year or a period of time by the estimated population of that year or period of time [13]. So, the formula for CDR is as follows:

$$CDR(y) = \frac{D(y)}{P(y)} \times 1000 \quad (1)$$

where:

$D(y)$ is the number of divorces per 1000 people in that year.

$P(y)$ is the number of people living in that year (population).

Using Equation 1, CDR (in the year 2017) in Saudi Arabia is 7.8 per 1000 people for both males and females. The CDR tends to be low since the age structure of the population and the gender male and female affect the CDR. So, the CDR has several limitations. It may not reflect changes in divorce accurately such as changes in age structure and marital status of population [13]. Therefore, due to these limitations, we used the General Divorce Rate (GDR) which is a refined measure compared to CDR. The GDR is the number of divorces in a given year or a period of time by the estimated population aged 15 and more in that year or period of time [13]. So, the formula for GDR is as follows:

$$GDR(y) = \frac{D(y)}{P(y)_{15+}} \times 1000 \quad (2)$$

where:

$D(y)$ is the number of divorces per 1000 people in that year.

$P(y)_{15+}$ is the number of people living in that year age 15 and older (population).

Using Equation 2, GDR (in the year 2017) in Saudi Arabia is 17.8 per 1000 people age 15 and older. We can conclude that there is a notable difference between $CDR = 7.8$ and $GDR = 17.8$ when we include only the people aged 15 and

more. We can also obtain the GDR separately for females and males by using the formula for General Divorce Rate by Gender (male/female) (GDRG) which is as follows:

$$GDRG(y) = \frac{D(y)}{P(y)_{15+}} \times 1000 \quad (3)$$

where:

$D(y)$ is the number of divorces for males or females per 1000 people in that year.

$P(y)_{15+}$ is the number of males or females living in that year age 15 and older (population). G represents the gender of the population.

Using Equation 3, the GDR for females GDR_f (in the year 2017) in Saudi Arabia is 24.1 per 1000 people age 15 and older. Similarly, mutatis mutandis, using this formula, the GDR for males GDR_m (in the year 2017) is 11.7 per 1000 people aged 15 and older.

From the previous calculations, we can represent the comparison between GDR for males and females age 15 and more in 2016-2017 as shown in Fig. 9. So, the GDR in Saudi Arabia for the people aged 15 and more at the year 2017 is 17.8 which is relatively high.

B. AGE-GROUP SPECIFIC ANALYSIS

Owing to our Saudi linked open data cloud, we were able to get more information related to population based on age structure. There is a strong relationship between divorce rates and ages [13]. We found that most of the divorced males are aged between ages 35 to 39, where divorced male population approximates to 15,000 and most of the divorced females are aged between ages 30 to 34, where the divorced female population approximates to more than 30,000. Therefore, we can summarize that most of the divorced people are aged between 30 to 39, whether male or female. This is followed by people aged 40 to 44, where the divorced male population is approximately more than 10,000 and the divorced female population numbers approximately more than 20,000. Therefore, it is also important to consider the variation in the age structure and gender of the population to measure the divorce rates by age group separately for males and females by restricting the measure to one gender and one age group at a time [13]. So, we can obtain the age-specific DR separately for females and males. Then the formula for age-specific divorce rate by gender (ADRG) is as follows:

$$ADRG_{(m,n)}^G = \frac{D_{(m,n)}^G}{P_{(m,n)}^G} \times 1000 \quad (4)$$

where:

$D_{(m,n)}^G$ is the number of divorces within a specific gender in the age group n to m in the year.

$P_{(m,n)}^G$ is the number of the people females/males population in the age group n to m in the year. G represents the gender of the population, G f,m.

The divorce rates are also affected by many other variables which are not directly predictable such as unemployment rates, educational status, income, household expenditure,

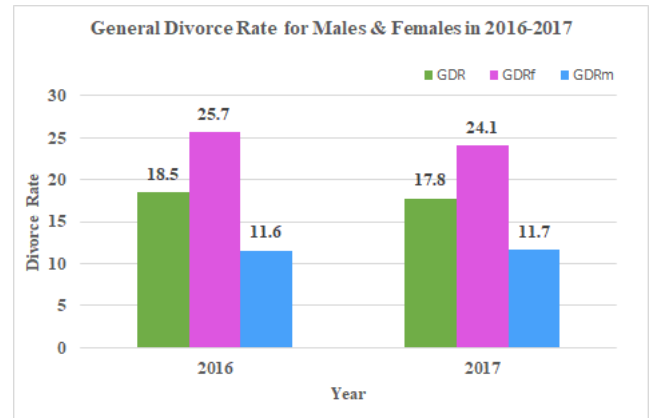


FIGURE 9. Comparison among GDR, GDRf and GDRm in 2016 and 2017.

consumption, and living costs [5], [15], [19], [29]. By using the Saudi linked open data cloud, we were able to find the relationship between the divorce rate and these variables and represent it mathematically using direct and inverse methods which is not possible without linking these different datasets together.

C. EDUCATIONAL-LEVEL BASED ANALYSIS

Our data elucidated that most of the divorced males have secondary (high school) education, where the divorced male population numbers approximately 30,000. Most of the divorced females have a university (bachelor) education, where the divorced female population numbers approximately 50,000. We can summarize that the divorce rates increase in people who have high school and bachelor education, whether they are male or female. The divorce rates decrease to the lowest rate in people who have Ph.D. education. We can also represent the DR by educational status for females and males from 2016 to 2017 as shown in Fig. 10. So, we can represent the relationship between divorce rates and educational status by using the inverse method. Since the DR decreases in people whether male or female who have higher educational status (ES) and vice versa, we can mathematically represent this relationship as follows:

$$C = DR(y) \times ES(y) \quad (5)$$

where:

C is a constant value do not change over time, it can be any type of number.

$DR(y)$ is the number of divorces in a given year by the estimated population aged 15 and more of that year.

$ES(y)$ is the number of people by the educational level for divorced males and females.

D. INCOME-BASED ANALYSIS

As discussed earlier, the divorce rates increase in people who have high school and bachelor education, whether they are males or females. Fig. 11 shows the average monthly wages for males and females by their educational level. The average monthly wages for males who have secondary

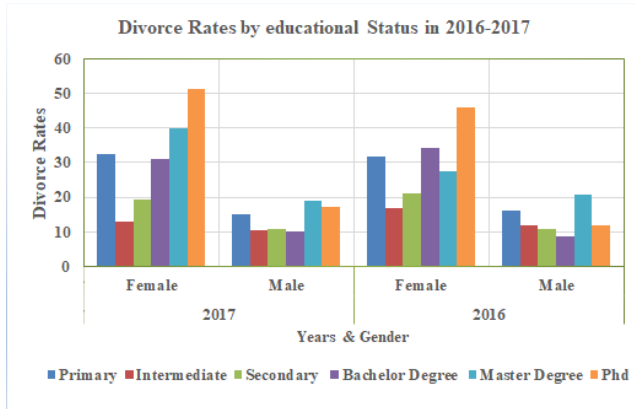


FIGURE 10. The divorce rates by educational status for females and males in 2016 and 2017.

(high school) education is approximately 10,000 RS and the average monthly wages for females who have a university (bachelor) education is approximately 10,000 RS. Also, the average monthly wages for males and females who have PhD education in which group the divorce rates is shown to decrease are between 20,000 to 25,000 RS. So, we can summarize that the people who lack financial stability tend to divorce while other people who have financially stability do not. In general, from what is mentioned earlier, we can also conclude that the income and educational levels affect the divorce rates in Saudi Arabia. Therefore, we can also represent the relationship between the DR and the average wages (W) for males and females by educational level by using inverse method since the DR increases in the people whether male or female who lack financial stability (low average wages) and vice versa. So, we can mathematically represent this relationship as follows:

$$C = DR(y) \times ES(y) \times W(y) \tag{6}$$

where:

W(y) is the average monthly wages for males and females by their educational level in a given year.

We can conclude that the divorce rates are affected by the educational level of people and their educational level in return affects their income so that the relationship between these three variables is inverse.

E. UNEMPLOYMENT-BASED ANALYSIS

The unemployment rate is also an important variable which affects divorce rates in Saudi Arabia. We have mentioned previously that the divorce rate increases in individuals who have high school and bachelor education, whether they are males or females. We found that the total unemployment rates for males and females are high for the people who have high school and bachelor education, wherein the total unemployment rate for males with high school is approximately 10.0, and the total unemployment rate of females with bachelor education is approximately 35.0. We can summarize that when the unemployment rates increase, the divorce rates also increase. As we mentioned earlier, the unemployment

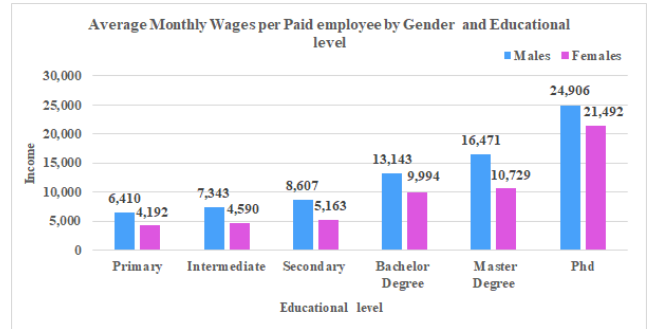


FIGURE 11. Income by educational status.

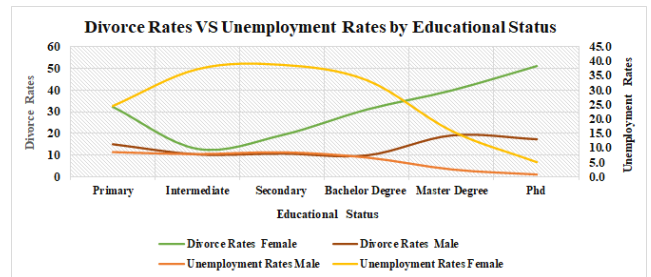


FIGURE 12. Comparison between the divorced rates and unemployment rates for females and males in 2017 by educational level in Saudi Arabia.

rates increase in males with high school education and the divorce rates also increase in males with high school education as shown in Fig. 12. In a similar manner, the unemployment rates increase in females with bachelor education and the divorce rates also increase in females with bachelor education. Therefore, we can also represent the relationship between the DR and the unemployment rates (UR) for males and females by educational level by using direct method since the DR increases when the UR increases and vice versa. So we can mathematically represent this relationship as follows:

$$C = \frac{DR(y)}{UR(y)} \times ES(y) \times W(y) \tag{7}$$

where:

UR(y) are the unemployment rates for males and females by their educational level in a given year.

VIII. SIMULATED VISUAL ENVIRONMENT FOR BROWSING AND ANALYZING SLODC

Visualization is the capability of presenting information for better understanding [14] by leveraging the users to understand datasets and simulation (in the context of semantic based applications) is more focused on importing and integrating data items in nodes from multiple query sources [18]. So, considering the semantic Web context, our end user interface simulated the data queried from different interlinked sources and visualized it in different formats including the iconic representation of data (as shown in Fig. 13). The ultimate background is to simulate and visualize data at three levels i.e. i) Ontology-RDF level, ii) graph relationship level, and data nodes level. Ontology-RDF level simulation and visualization involves simulating and integrating the concepts



FIGURE 13. Simulated visual environment of SLODC, showing divorce rate in Saudi Arabia based on different factors.

and relationships between concepts. Graph relationship level simulation and visualization involves identifying data entities in datasets, links between data entities and simulated query results from different datasets. To get the benefits of first two levels of simulation and visualization, the linked open data consumers should have pre-knowledge of the dataset’s data model and SPARQL queries. Therefore, in the data visualization type, linked open data transformed into rich visuals provides better decision making. It also involves visualizing data with different types of charts, maps, and graphs to present different visual means for performing data analysis.

Our simulated visual environment is interactive in many ways such as it can be used to visualize, navigate, explore, and query the SLODC (as shown in Fig. 13). This environment is created by dumping the RDF datasets in the GraphDB and simulating these live datasets with visual environment (i.e., SPARQL endpoint) by using power APIs and OWL APIs. The OWL APIs are used to get query results, simulate them, and visualize them by using Power APIs. Saudi linked open government data consumers such as researchers, journalists, decision-makers, and activists can benefit from such data. However, they cannot do so because they lack expertise and technical knowledge related to data management and visualizations. The interactive and easy to use interface dashboard of Saudi linked open government data helps to have a better understanding and communication with large amounts of data.

To create a simulated visual environment, we followed the following process; we extracted SPARQL results after querying Saudi linked open government data to be prepared for the visualization, we pre-processed the extracted results transformed, and modeled before they can be visually analyzed. Then, after extracted results are processed, we visualized

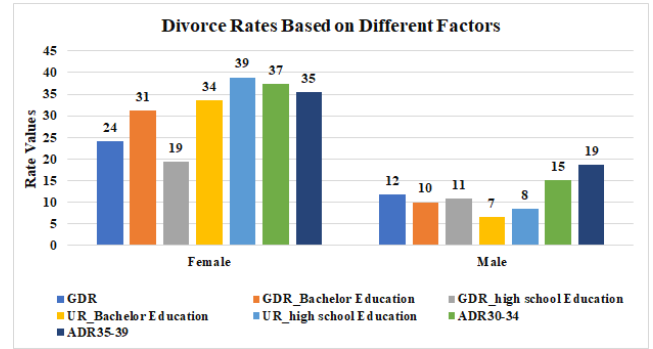


FIGURE 14. Divorce rates based on different factors in Saudi Arabia.

using different types of charts, enriched and customized with visual characteristics, and aggregations. Finally, we developed interactive dashboards that use visualizations to tell a story, provide interactive analysis, decision making, and insights. An important feature of our simulated visual environment is that we used semantic matching of icons with the related results (as shown in Fig. 13) for better understanding and analyzing the results.

IX. POLICY MAKING FOR BENEFIT OF SOCIETY

Once, the SLODC is created and a simulated visual environment is configured on top of this data cloud, it can be used to browse, navigate, query, and visualize the statistics of interlinked data by different stakeholders including policy and decision makers. As an example, the results (as shown in Fig. 14) can be used by policymakers to define future policies to control the increase in divorce rates. In Fig. 14, it describes the divorce rates based on different factors such as General Divorce Rates (GDR) for Males and Females. The General Divorce Rates (GDR) in people (Males/Females) with high school and bachelor education where the divorced male population numbers approximately 30,000 and the divorced female population numbers approximately 50,000. We can conclude that the divorce rates are affected by the educational level of people and their educational level in return affects their income, so that the relationship between these three variables is inverse. Thus, we can conclude that the people who lack financial stability tend to divorce while other people who are financially stable do not divorce. Moreover, the Age Group-specific Divorce Rate (ADR) for people (Males/Females) aged 30 to 39, where the divorce rate for males age interval 35 to 39 is 18.6 and the divorce rate for females age interval 30 to 34 is 37.3. The Unemployment Rate (UR) in people (Males/Females) with high school and bachelor education is as follows: the total UR for males with high school education is approximately 10.0 and the total UR for females with bachelor education is approximately 35.0. Also, the household expenditure, consumption, and living costs are directly affected by the divorce rates in Saudi Arabia, and the relationship between the divorce rates and other variables the educational level of people and their income is inverse.

Based on the SLOCD analysis, efficient, sustainable, and effective policies can be developed by policy and decision makers to address different issues and factors that cause the increase in divorce rates. Some of the factors that we analyzed in above discussion are, low education rate, which results into low income, which causes low standard lifestyle, which ultimately results to divorce. So, it means if governments define policies to support the people at all these three levels (with same sequence of priorities), divorce rate could be controlled. Some of the policies that can be defined to reduce divorce rates may include efforts to promote programs that emphasize the importance of marriage, stable and strong family life and relationships, provide counselling services as well as advice and help to people before and after the marriage. Also, offering material support for married couples and divorced people may also be helpful. This can include establishing a counselling office for providing advice to reduce divorce in Saudi society and to provide prospective spouses and couples with access to social and psychological specialists. This can help to ensure the compatibility between the spouses and provide advice to those wishing to marry so that they can enjoy a more stable and harmonious marital life. The importance of family life and the negative consequences of divorce for individual, family, and society can also be addressed in educational courses at school. Viewing as limitation of the system, such policies (based on data analysis) has been prepared and proposed to higher authorities but not yet implemented. Also, our system needs to be continuously updated with latest data from different sources to get right results from right data. For this purpose, system needs to be maintained with latest datasets.

X. CONCLUSION

Open government data have been proven to be an area of great challenge for the Saudi government since it launched the initiative of developing the open data portal. This study presented an extended framework wherein linked open data technologies are used to create Saudi linked open data cloud by retrieving and interlinking related data entities from various and heterogeneous data sources. Additionally, we also presented a comprehensive analysis to explore the factors impacting the rapid change in social values in Saudi Arabia. This would not have been possible without using linked open data technologies. More specifically, we built a Saudi linked open government data cloud which helped to identify the relation between different factors which were not linked before and how they relate and affect the different social attributes such as increasing divorce rate in Saudi Arabia. We extracted and produced RDF data from various domains such as population, household income, educational status, job opportunities, household expenditure, consumption and living costs to create Saudi linked open data cloud of these social data entities. We established SPARQL endpoint on top of this cloud and posed various queries to retrieve data from the Saudi linked open government data cloud. To investigate the computational impact of our work on society, and as a practical case study of our work, We investigated the

impact of different factors and variables on rapidly increasing divorce rates which is a challenging social issue in Saudi Arabia. Our quantitative and qualitative analysis of Saudi linked open government data cloud revealed that different factors contribute towards the increasing number of divorce rates in the region. Measuring the system (framework) performance is an important aspect that could help in adopting the framework by concerned stakeholders. System performance can be measured by using various factors such as accuracy, time, and GUI usability. However, depending on the nature of our current study, time/speed does not seem a realistic measure. So, to measure the system performance, we focused more on data accuracy and its visual presentation to decision and policy makers in an easy-to-understand way. Data accuracy is measured by performing hundreds of queries related to our case as well as to the data which are part of LOD cloud by team of researchers having good hands on the SPARQL query. The results of SPARQL queries are cross verified with source data to measure the data retrieval accuracy. Similarly, the results are presented to domain experts in visual/graphical forms through the dashboard for their interpretation and understanding which is cross verified by the IT experts. It is expected that the findings of such studies can be used by decision and policy makers to define national policies to strengthen and value the social relations in society.

The presented work can be extended in numerous ways. One potential extension could be to combine the linked open data technologies with (deep) machine learning for more informed reasoning and decision making.

REFERENCES

- [1] M. A. Smadi, I. A. Dalabih, Y. Jararweh, and P. Juola, "Leveraging linked open data to automatically answer Arabic questions," *IEEE Access*, vol. 7, pp. 177122–177136, 2019.
- [2] C. Alexopoulos, E. Loukis, S. Mouzakitis, M. Petychakis, and Y. Charalabidis, "Analysing the characteristics of open government data sources in Greece," *J. Knowl. Economy*, vol. 9, no. 3, pp. 721–753, Sep. 2018.
- [3] A. M. Alsukhayri, M. A. Aslam, S. Arafat, and N. R. Aljohani, "Leveraging the Saudi linked open government data: A framework and potential benefits," *Int. J. Mod. Educ. Comput. Sci.*, vol. 11, no. 7, pp. 14–22, Jul. 2019.
- [4] A. M. Alsukhayri, M. A. Aslam, K. Saeedi, and M. S. A. Malik, "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*, vol. 12, no. 20, p. 8608, Oct. 2020.
- [5] L. Anser, "Divorce in the Arab Gulf countries," in *Contemporary Issues in Family Studies: Global Perspectives on Partnerships, Parenting and Support in a Changing World*, 2014, pp. 59–73.
- [6] P. R. Aryan, F. J. Ekaputra, W. D. Sunindyo, and S. Akbar, "Fostering government transparency and public participation through linked open government data: Case study: Indonesian public information service," in *Proc. Int. Conf. Data Softw. Eng. (ICODSE)*, Nov. 2014, pp. 1–6.
- [7] M. A. Aslam and N. R. Aljohani, "SPedia: A central hub for the linked open data of scientific publications," *Int. J. Semantic Web Inf. Syst.*, vol. 13, no. 1, pp. 128–147, 2017.
- [8] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "DBpedia: A nucleus for a web of open data," in *The Semantic Web*. Berlin, Germany: Springer, 2007, pp. 722–735.
- [9] L. P. Batagan, D.-L. Constantin, and L. M. Moga, "Facts and prospects of open government data use. A case study in Romania," in *Citizen Empowerment and Innovation in the Data-Rich City*. Cham, Switzerland: Springer, 2017, pp. 195–208.

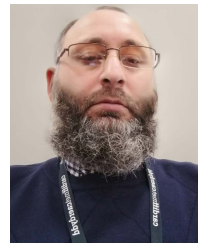
- [10] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data: The story so far," in *Semantic Services, Interoperability and Web Applications: Emerging Concept*. Pennsylvania, PA, USA: IGI Global, 2011, pp. 205–227.
- [11] L. Boonlamp, "A linked data approach to planning collaboration amongst local governments in Thailand," in *Proc. 2nd Int. Conf. Inf. Technol. (INCIIT)*, Nov. 2017, pp. 1–5.
- [12] T. Chen, Y. Zhang, Z. Wang, D. Wang, H. Li, and W. Liu, "SinoPedia—A linked data services platform for decentralized knowledge base," *PLoS ONE*, vol. 14, no. 8, Aug. 2019, Art. no. e0219992.
- [13] P. Chhaya, C.-H. Choi, K.-H. Lee, W.-S. Cho, and Y.-S. Lee, "KMLoD: Linked open data service for Korean medical database," *J. Supercomput.*, vol. 76, pp. 7758–7776, Oct. 2017.
- [14] A. R. Diaz, A. Benito-Santos, A. Dorn, Y. Abgaz, E. Wandl-Vogt, and R. Theron, "Intuitive ontology-based SPARQL queries for RDF data exploration," *IEEE Access*, vol. 7, pp. 156272–156286, 2019.
- [15] P. Dommaraju and G. Jones, "Divorce trends in Asia," *Asian J. Social Sci.*, vol. 39, no. 6, pp. 725–750, 2011.
- [16] A. Hogan, *The Web of Data*. Cham, Switzerland: Springer, 2020.
- [17] A. Hosen and I. Alfina, "Aggregation of open data information using linked data: Case study education and job vacancy data in Jakarta," in *Proc. Int. Conf. Adv. Comput. Sci. Inf. Syst. (ICACSIS)*, Oct. 2016, pp. 579–584.
- [18] Q. Huang, J. Zhou, L. Shi, and W. Yan, "Simulation of a semantic web data sharing system based on NS-3," in *Proc. IEEE/ACIS 13th Int. Conf. Comput. Inf. Sci. (ICIS)*, Jun. 2014, pp. 409–414.
- [19] A. Isen and B. Stevenson, "Women's education and family behavior: Trends in marriage, divorce and fertility," Nat. Bur. Econ. Res., Cambridge, MA, USA, Tech. Rep. 15725, 2010.
- [20] G. Jaglan and S. K. Malik, "LOD: Linking and querying shared data on web," in *Proc. 8th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence)*, Jan. 2018, pp. 568–573.
- [21] M. Jovanovik, M. Petrov, B. Najdenov, and D. Trajanov, "Linked music data from global music charts," in *Proc. 10th Int. Conf. Semantic Syst.*, 2014, pp. 108–115.
- [22] M. Jovanovik and D. Trajanov, "Consolidating drug data on a global scale using linked data," *J. Biomed. Semantics*, vol. 8, no. 1, pp. 1–24, Dec. 2017.
- [23] P. Krataithong, M. Buranarach, and T. Supnithi, "RDF dataset management framework for data.go.th," in *Proc. Int. Conf. Knowl., Inf., Creativity Support Syst.* Cham, Switzerland: Springer, 2015, pp. 42–53.
- [24] B. MacKellar, C. Schweikert, and S. A. Chun, "Patient-centered clinical trials decision support using linked open data," in *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications*. Cambridge, MA, USA: IGI Global, 2018, pp. 1765–1782.
- [25] R. M. Nawi, S. A. M. Noah, and L. Q. Zakaria, "Integration of linked open data in collaborative group recommender systems," *IEEE Access*, vol. 9, pp. 150753–150767, 2021.
- [26] S. Mouzakitis, D. Pappaspyros, M. Petychakis, S. Koussouris, A. Zafeiropoulos, E. Fotopoulou, L. Farid, F. Orlandi, J. Attard, and J. Psarras, "Challenges and opportunities in renovating public sector information by enabling linked data and analytics," *Inf. Syst. Frontiers*, vol. 19, no. 2, pp. 321–336, Apr. 2017.
- [27] B. Šlibar, D. Oreški, and N. B. Redep, "Importance of the open data assessment: An insight into the (meta) data quality dimensions," *SAGE Open*, vol. 11, no. 2, Apr. 2021, Art. no. 215824402110231.
- [28] M. N. Spojala, O. Rinciog, and V. Posea, "The semantic representation of open data regarding the Romanian companies," in *Proc. 15th RoEduNet Conf., Netw. Educ. Res.*, Sep. 2016, pp. 1–5.
- [29] T. E. Trail and B. R. Karney, "What's (not) wrong with low-income marriages," *J. Marriage Family*, vol. 74, no. 3, pp. 413–427, 2012.
- [30] B. Villazón-Terrazas, L. M. Vilches-Blázquez, O. Corcho, and A. Gómez-Pérez, "Methodological guidelines for publishing government linked data," in *Linking Government Data*. New York, NY, USA: Springer, 2011, pp. 27–49.
- [31] P. Yochum, L. Chang, T. Gu, and M. Zhu, "Linked open data in location-based recommendation system on tourism domain: A survey," *IEEE Access*, vol. 8, pp. 16409–16439, 2020.

AFNAN MUHAMMADSALH ALSUKHAYRI received the bachelor's and master's degrees in information technology from the Department of Information Systems, King Abdulaziz University, Jeddah, Saudi Arabia, in 2021. She is currently a Data Scientist with the Web Observatory Center for Web Studies, King Abdulaziz University. Her research interests include linked open data, semantic web, and social network analysis.



MUHAMMAD AHTISHAM ASLAM received the master's degree in computer science from Hamdard University, Lahore, Pakistan, in 2002, and the Ph.D. degree from the University of Leipzig, Leipzig, Germany, in 2007. From 2003 to 2007, he was a Research Assistant at the University of Leipzig. He has been working as a Senior Staff Researcher at the Intelligence Laboratory, Knowledge Technology Cluster, Malaysian Institute of Microelectronic Systems (MIMOS),

Kuala Lumpur, Malaysia. He also has been working as an Assistant Professor at the COMSATS Institute of Information Technology, Lahore. He is currently working as an Associate Professor with the Department of Information Systems, King Abdulaziz University (KAU), Jeddah, Saudi Arabia. His research interests include semantic web, linked open data, data science, and social networks. He was a recipient of the 13th All Pakistan Dr. Abdul Qadir Khan Research Laboratories Software Competition Award, in 2002, the Partial Support Scholarship for Ph.D. Studies Abroad from the Higher Education Commission (HEC) of Pakistan, the Best Researcher Award from the Department of Information Systems, Faculty of Computing and Information Technology (FCIT), KAU, for the year 2016–2017, and the Best Session Research Paper Awards in 8th COMSATS Workshop on Research in Computing (CWRC 08), Wah Cantt., Pakistan.



IMTIAZ HUSSAIN KHAN received the M.S. degree in computer science from the University of Essex, U.K., in 2005, and the Ph.D. degree in natural language generation from the University of Aberdeen, U.K., in 2010. He is currently working as a Senior Lecturer with the School of Technologies, Cardiff Metropolitan University, U.K. He has published more than 30 articles in well-reputed journals, including *Cognitive Computation* and *IEEE Access*. His research interests include natural language processing and evolutionary computation.



RABEEH AYAZ ABBASI received the Ph.D. degree from the Institute of Web Science and Technologies (WeST), University of Koblenz-Landau, Germany. He is a Tenured Associate Professor with the Department of Computer Science, Quaid-i-Azam University, Islamabad, Pakistan. He has a vast research experience in the fields of social media analytics and social network analysis. He has published more than 50 articles in reputed journals, including *IEEE Computational Intelligence Magazine*, *Computers in Human Behavior*, *Telematics and Informatics*, *Applied Soft Computing*, *Knowledge-Based Systems*, *Journal of Informetrics*, *Expert Systems with Applications*, and *Scientometrics* and international conferences including, ACM HyperText Conference (HT), ACM World Wide Web Conference (WWW), Pacific Asia Conference on Knowledge Discovery and Data mining (PAKDD), and European Conference on Information Retrieval (ECIR). His research interests include leveraging positive aspects of social media, including social media's use in saving lives, understanding events, and analyzing sentiments among many others.

AMAL BABOUR received the Ph.D. degree in computer science from Kent State University, Kent, OH, USA, in 2017. She worked as a Visiting Assistant Research Professor at the Networking and Media Communications Research Laboratory, Computer Science Department, Kent State University. She is currently an Assistant Professor with the Information System Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. She has participated as a Reviewer for multiple journals and conferences, such as *JMIR*, *Applied Sciences*, *IEEE Access*, *ICISDM*, and *IHCI*. Her research interests include artificial intelligence, machine learning, crowdsourcing, and information retrieval.

• • •