Leveraging the Saudi Linked Open Government Data: A Framework and Potential Benefits

Afnan M. AlSukhayri, Muhammad Ahtisham Aslam, Sachi Arafat and Naif Radi Aljohani
Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
Email: {aalsukhayri, maaslam, sarafat, nraljohani}@kau.edu.sa

Abstract—Open data initiatives are a crucial aspect of effective e-governance strategy. They embody aspirations towards sociopolitical values of transparency, trust, confidence, and accountability, pertaining to the relationship between a government and its citizens. The importance of such initiatives is especially important for an emerging economy such as Saudi Arabia which is undergoing rapid social changes directed by a contemporary national vision. The effectiveness of open data initiatives depends strongly on (a) the quality of the data available, (b) the soundness of the methodologies and suitability of platforms used to prepare and present the data, and (c) the ability of the data to facilitate the kinds of insights and social-action that are sought from that data to ensure successful e-governance. This paper investigates the feasibility of current Saudi government open data initiatives in this regard. It assesses existing approaches to improve the effectiveness of open government data through transforming it into linked-open data (using the Resource Description Framework [RDF]) by connecting disparate sources of structured data therein. It proposes to improve existing approaches by suggesting a framework for automating the linking sub-process of existing approaches and organizing the data to be queried through SPARQL. Moreover, it evaluates the potential benefit of this proposal by discussing the kinds of policy insights this could generate which would be difficult without it.

Index Terms—Open data, knowledge engineering, smart query, government data, Linked Open Data, machine reasoning, policy making, data science.

I. INTRODUCTION

Linked Open Data (LOD) refers to a set of practices to publish, share and connect data on the Web in such a way that it is machine-readable [1]. In recent years, many governments have used LOD to create platforms, portals, services, and applications as an effective form of Open Government Data (OGD). Its effectiveness here stems from the accompanying metadata and ontologies that help to link disparate data items, help make more sense of the data, and facilitate the generation of more insights using analytical tools.

Governments support open data as a means to improve e-governance strategies such the Linked Open Government Data (LOGD) initiatives in the US, UK, and Singapore [1-3] that have resulted in increased government transparency, public participation and access to government data. They have enabled the creation of value by providing access to and querying of multiple data sources. Open data initiatives are a means for sociopolitical and urban development of smart cities [4].

Saudi Arabia is a country with a rapidly changing socio-political landscape. Its policies increasingly depend on effective e-governance to ensure the fulfillment of the typical values of government that are of particular importance to its current states such as transparency, trust, accountability, and openness. To this end the Saudi government have developed an Open Data portal and related legal policy to obligate or encourage government and other institutions, to simplify citizen access to data. These are not isolated policies but central to their 2030 vision [5].

The existing portal however, is quite limited. The data is stored in distributed databases on different sub-portals and documents, and suffers from duplication and inconsistency. Links between different data entities with related content are often missing. Moreover, the data is often not available in a machine-processable format or with metadata sufficient to allow for meaningful analytics.

The framework we propose here seeks to remedy some of these problems. In particular, we propose a system to automate aspects of the process to generate LOGD from unlinked but structured data in Saudi government portals. This paper presents the architecture of our proposed framework and describes the function of its component modules. It explains the algorithm developed to produce open data from different data sources. We suspect that an equivalent algorithm has yet to be developed in other (i.e., non-Saudi) LOGD contexts. While there are extensive open government datasets, they are often limited to manual interlinking of datasets and do not provide query facilities such as through SPARQL endpoints [6].

Our proposed framework focuses on bridging the gap between public and government sectors by producing Saudi OGD and making a SPARQL endpoint available for intelligent question answer purposes. The remainder of the paper is organized as follows: Section II describes

the motivation for this work. Related work is discussed in Section III. Section IV describes our proposed framework, the Linked Open Saudi Government Data Framework (LOSGDF). In section V, we describe the potential benefits of producing and using LOGD by performing test queries via a SPARQL endpoint on resulting open datasets. Finally, Section VI concludes our work giving directions for future work.

II. MOTIVATIONS

Open data is important for modernity for several reasons. It is a necessary step towards fulfilling several values pertaining to successful e-governance including transparency, accountability, trust, and confidence. The importance of it to fulfill any of these values is hotly debated and depends on the specific country and its setup [7], and the factors that work to amplify or reduce its significance [8]. Its importance is not limited to a top-down strategy of governance but are tied to (a) to bottom-up governance models that value civic engagement and participatory governance [8], and to (b) various socio-political or socioeconomic goals that may not explicitly refer to governance [7,9-11], such as disaster recovery [12], reduction of public expenditure and improvement of the economy [13, 14].

In a country such as Saudi Arabia with a rapidly changing socio-political landscape, effective e-governance is a necessity for progress in many areas. The government has issued many legal policies to obligate or encourage government institutions or otherwise, to facilitate access to data for its citizens. These policies form a core aspect of an unprecedented national vision, the 2030 vision [5]. Legislation is an important step towards developing effective e-governance. Successful e-governance is also however dependent on other factors such as (i) institutional setup [15], (ii) policies related to these values that are not explicitly about open-data [7, 8] and (iii) the effectiveness of the data and tools available for making use of that data for decision making (from the citizen's or government's viewpoint) specifically and civic engagement in general. We focus on this third aspect while acknowledging that all three groups of factors are intricately linked such that successful fulfillment of the respective goals requires progress in all three areas in an integrated and parallel fashion.

An increasing number of governments have taken solid technological steps pertaining to (iii) (and steps corresponding to (i-ii) that are ancillary) to create Open Government Data initiatives [13-16, 20-21]. However, despite their numerous efforts and heavy investments in publishing and consuming OGD, its impact and benefits in most national contexts have yet to materialize at the level of organizations and individual citizens. Thus, the values such initiatives are purported to engender remain unrealized. There are several reasons for this, but they can be roughly summarized into two. First, while being freely available is a key benefit of open data, this cannot be actualized when the data is of bad quality (e.g. inaccurate, unreliable or non-actual data, untimely, not

sufficiently detailed), badly formatted or in formats that are not open (and hence require paid software to access) way, or presented in an insufficiently organized way. Secondly, beyond this first reason, the simple availability of numerous such databases without any facility for the average citizen or organization to easily generate insights from them, is not sufficient to realize the values mentioned above.

The first set of problems is important. Other than the usual automated techniques for detecting and resolving bad quality data these problems often require resolution through lobbying the relevant authorities. The second set of problems depend on improving methods and technologies for optimal data preparation (for data that is already of good quality) so that the maximum number of insights can be generated from it. We focus on this second set of problems; they require the depth and breadth of scientific and computational approaches to solve. The following section details some of these approaches.

III. RELATED WORK

Open data initiatives are increasingly common and in addition to making data available, governments often also provide supporting platforms/portals and analytics tools, services and related applications [16]. LOD is an active area of research in industry and academia, and involves anything from the development of algorithms for automatic linking, to sociotechnical theorization about LOD based business and social processes. Like the government initiatives, academic LOD research has produced several frameworks and tools for extracting, producing and publishing LOD [17]. This includes the establishing of SPARQL endpoints for query answering, web applications for data analytics and visualization, and tools for linking of government datasets to those in the open data portal [17]. There are of course many examples of LOGD for various countries, with research elucidating the corresponding platforms and assumptions [7-11]. However, it is their particular limitations when applied to our problem context that motivated our work.

To this end the approach in [17] is limited by its lack of information about organizations published in sources such as dbpedia.org and lack of adequate vocabulary for the description of properties. The LOD solution by the government of Greece in [18, 19] is another prominent example of LOGD. Moreover, it can be used to search integrated information from various information sources as a large knowledge graph. The tool in [18] facilitates the storage, management, manipulation of RDF datasets and can also be used for querying these datasets by using SPARQL. It can also be integrated with the ERMIS Greek portal for public administration to transform data into LOD. However, this requires the English version of web pages. In addition, they failed to assess their approach in extended datasets and by comparison with equivalent web pages in European countries. In [19], the e-GIF ontology was applied to the Greek OGD in order to associate it to external datasets such as DBpedia by first

transforming web pages into RDF. They propose a solution that uses the JENA framework, the D2RQ language, and Silk framework to produce LOD from the RDF triples. Their approach suffers from major issues: a) it needs to download web pages to create a local copy, b) the e-GIF ontology needs to cover all types of concepts and relations to accurately describe web pages and c) the appropriate choice of tools to produce LOD is unclear.

The work in [23] propose a framework to link and query open datasets from various government portals and translate the data to graph based on RDF data. An online portal was used as a source for open data but the individual data entities therein were not interlinked. A framework for LOGD was developed in [24] for the government of Thailand. It uses D2R technology for publishing the RDB to RDF on the semantic web which allows for the browsing and searching of RDF data with SPARQL. The framework was further evaluated using road construction data and found to be effective in facilitating access, discovery and efficient in operation. However, the result presentation site is difficult to use, it requires data to be programmed.

The work in [25] combines data from different organizations into RDF format, links them and provides a GIS platform for visualizing the data. The approach therein facilitates the answering of identified competency questions. Its associated applications employ data from various sources which it integrates programmatically making the resultant information visually available to users. The work validates its overall approach through a case-study on flooding in the Rio Doce Basin in Brazil using SPARQL.

The main relevant work employing Saudi LOGD as we intend to be [22]. The usefulness of the data therein was investigated through the issuance of numerous queries on data pertaining to educational statistics. Since the data is limited to the educational domain the assessment of the impact of linking heterogeneous domains on such government data is not immediately possible. We were motivated by this shortcoming, and sought to engage in rigorous analysis based on linked-domains. Moreover, there were several other shortcomings of [22] to do with its storage requirements and SPARQL access limitations that we are also keen to improve-upon. Our main aim however is to create a system to further automate the semi-automatic process [22] for transforming data into RDF.

Table I. Cooperative studies between different studies according to TOOLS AND techniques in logd.

Tools/ Techniques	Greek LOD	LSD	Romania LOGD	Indonesia LOGD	LGD
Jena Framework	√	√		√	
D2R Framework					√
Triple Store	√	√			√
SPARQL	√	√	√	√	
Visualization Tools			√		

Table I summarizes the features of the solutions presented in the discussed works. Most studies such as that on Greek Linked Open Data, Linked spending data, and the Indonesian LOGD used SPARQL endpoints and the Jena framework. We therefore took these technologies as relevant for investigation for our own project. There are of course a variation of technologies that are used. The Thai local government employed SPARQL endpoints and the D2R Framework rather than the Jena framework to publish their data into RDF. The Romanian government instead paired SPARQL endpoints with visualization tools. Our proposed approach seeks to automate the metadata generation and linking process pertaining to structured government data on the online portal (e.g. data.gov.ro).

IV. LOSGDF FRAMEWORK

This section describes our proposed framework that we propose for automated metadata generation and linking. Its overall purpose to automate the process of creating LOGD data from previously unlinked structured data. In particular, this framework which we call Linked Open Saudi Government Data Framework (LOSGDF) is for collecting, processing, generating RDF datasets, interlinking these datasets with other open datasets and store them into a triple store server. The overall architecture of LOSGDF Framework consists of four modules, the data preparation, modeling, linking and querying modules, as depicted in Fig.1. The following section describes each of these in turn.
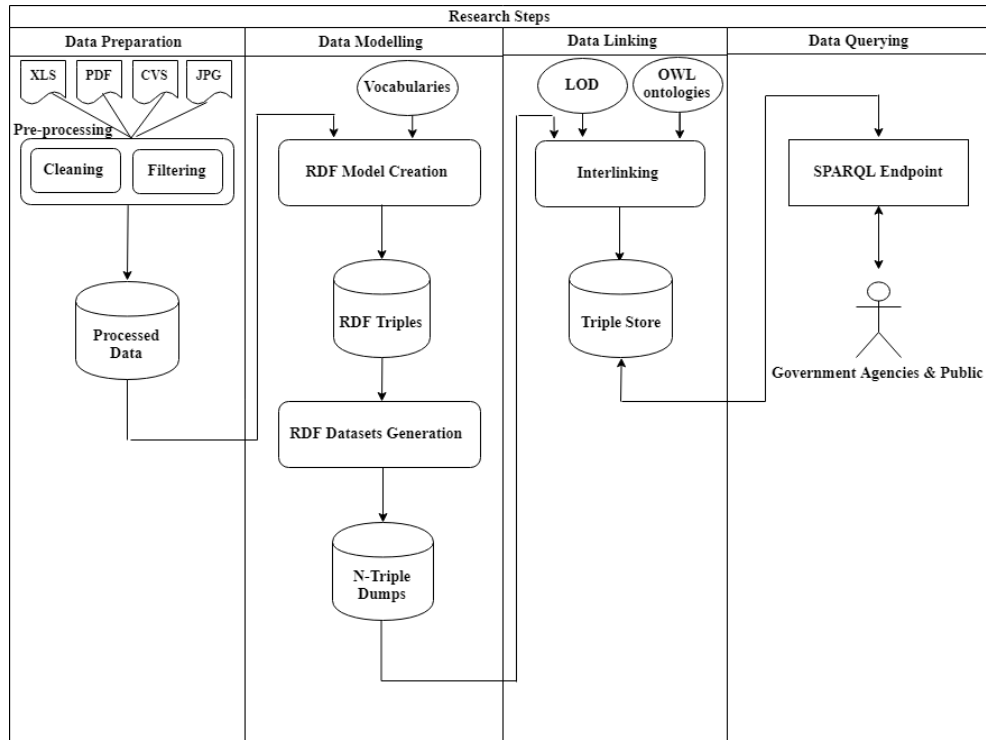
Fig.1. Architecture of LOSGDF framework.

A. Data Preparation

The data preparation module consists of the data collection and pre-processing sub-modules. The data collection module is responsible for traversing different data sources such as websites and government portals to collect relevant data entities. Our framework focuses on the open data available on Saudi ministry websites and portals. Data collection will also involve following links between data objects such that connected data is also collected. Linking open datasets involves investigating relations between government organizations (between Open Government datasets) supervised by certain ministries, establishing and then publishing them. This process of finding, extracting, linking and publishing the interlinked datasets follows the algorithm in Algorithm 1.

Algorithm 1. The algorithm of collecting Open Government Data (OGD)

```
Data: Saudi Ministries Websites/portals as data source
Result: Open Government Data in different format (i.e. CVS,
XLS, DBs etc.)
While Not traverse the whole Ministry website/portal do
   Traverse Ministry website/portal;
   Search for published OGD;
   Collect OGD;
   If There is a link to other government website then
      Follow link;
      While Not traverse the whole government, website do
         Traverse government organization website;
         Search for published OGD;
         Collect OGD;
      End
   End
End
```

The second sub-module is the data pre-processing module. It performs its job in two steps, data cleaning and data filtering. Each dataset collected from previous step is subject to pre-processing to address data heterogeneity. The collected datasets exist in excel sheet, databases, and CVS forms. Therefore, the datasets must be clean so that its relevant aspects can be extracted, filtered and transformed into a structured form. This process results in clean data that is ready for modelling by the next module.

B. Data Modelling

Data modeling of extracted data takes place in two stages, RDF model creation and RDF dataset generation. In order to create the RDF data model, we investigated different existing vocabularies that can possibly be reused or extended to model the data processed by the previous module. Once the data model is finalized, the data would be transformed into RDF format (*Subject - Predicate – Object*) with chosen vocabularies using a tool such as Google Refine [23]. Table II shows an example dataset of the Female population by administrative Areas and marital status available in excel sheet published in Saudi Open Data portal by the Central Department of Statistics and Information. Table III shows a sample of RDF statements belongs to a dataset in Table II generated by using our LOSGDF framework. This step would result in RDF triples. In the second step (i.e. RDF datasets generation), the RDF triples created from the previous step will be used to generate RDF datasets in N-triple format.

Table II. Sample of datasets published in Saudi open data portal.

Administrative Area	Marital Status			
	Married	Divorced	Widowed	Never Married
Al-Riyadh	891118	35628	69432	462273
Makkah	895433	49789	114262	547723
Al-Madinah	271384	10375	32238	150216
Al-Qaseem	196730	6852	22945	122082
Eastern Region	657146	19404	54419	326352
Aseer	364081	18967	45122	201882
Tabouk	142286	5450	14048	74167
Hail	111919	4600	15735	62765
Northern Borders	54287	2288	6284	28903
Jazan	285394	7749	28044	147111
Najran	95455	2720	10195	40953
Al-Baha	96883	1638	9452	38050
Al-Jouf	70547	2603	6899	35506

Table III. Sample of RDF statements belongs to a dataset in table ii generated by using LOGD framework.

Subject	Predicate	Object
SOD: Al-Riyadh	SOD:has_married_total	891118
SOD: Makkah	SOD:has_Widowed _total	114262
SOD: Najran	SOD:has_Divorced_total	2720

C. Data Linking

Once a data model is created and RDF datasets generated, data entities would require to be linked with each other both within the dataset and to other available open datasets so as to realize Linked Open Government Data (LOGD). In addition to linking open government datasets, this module would also require to identify and choose some external datasets to be linked to our Open Government datasets. Wikipedia [26] could be used for its textual content about, for example, events [26], places and entities in Saudi Arabia. Geonames [1] could be used as it includes spatial data, World Bank [1] and CIA Factbook datasets for economic and demographic information on Saudi Arabia and related countries. In order to do this, we first identify the entities close to the real world as classes, subclasses, objects, and properties. Then, after choosing the potential external datasets, we identify the common classes between datasets and relationship between objects from different datasets. This can be done by using a tool such as SILK [19] to automate the interlinking process between datasets using

OWL properties such as owl:sameAs and other properties based on similarity measures for example:
<http://www.semanticweb.org/SaudiOGD#Al-Riyadh> <owl:sameAs> <http://www.dbpedia.org/resource/Riyadh_Region>

Linking different datasets can lead to the access of more data entities and help retrieve more information that can ultimately result in a bigger knowledge graph from multiple interlinked datasets.

D. Data Querying

This module plays the role of creating in a SPARQL endpoint for the final extracted and produced linked open government datasets. These RDF datasets would be saved in the Triple Store Server and a public access point made available. The SPARQL endpoint can be used to ask smart questions by making use of the SPARQL protocol. This is not otherwise possible using traditional data. At this stage, we can perform intelligent querying to generated RDF datasets to evaluate the quality of the data stored in the RDF Triple Store Server. The next section describes some SPARQL endpoints and the results of some queries over this SPARQL endpoint. It also presents the quantitative and qualitative analysis of the resultant data, and shows how publishing LOGD can be used to make integrated queries between organizations, government agencies, and public datasets.

V. Results and Data Analysis

The open data initiatives and in particular LOD is a growing movement for government to make their data available in a machine-readable format [1] and its impact is particularly being felt in emerging economies [3]. Different governments around the world have already taken initiatives to publish their public data as open data and to link it the existing data sources such as the Linked Open Data Cloud [27]. The LOD cloud contains 1,239 datasets with 16,147 links, in domains such as media, biology, chemistry and economics [27]. DBpedia being one of the biggest datasets in the LOD Cloud is considered a hub of this Web of Data. DBpedia datasets are extracted from Wikipedia articles [26] and have a bidirectional link to many datasets in the LOD Cloud. Keeping in view the value of LOD and SPARQL queries, we present here some sample queries made to a DBpedia[1] SPARQL endpoint and some queries over an EU linked open data[2] SPARQL endpoint. The goal is to show the value of producing and linking open data, and to describe the needs of frameworks that can produce such data. We also show the results of these queries and analyze the results in order to explore the benefit of LOD.

A. Queries generated from DBpedia datasets

DBpedia contains structured information from Wikipedia that is available on the Web. DBpedia provides real data spanning various domains. It allows us

---

[1] http://dbpedia.org/sparql
[2] http://data.europa.eu/euodp/en/linked-data

to ask queries against Wikipedia, and to link with other datasets on the Web.

In this section, we report on the queries we issued to DBpedia SPARQL endpoint to explore government data. We identified different demographics factors that play a key role in economic growth. DBpedia queries pertaining to the Saudi economy and demographics lead us to insights how demographics drive the economy. The Saudi economy is dependent on oil and the government has a strong influence on major economic activities. Economic growth depends on gains in productivity and that is dependent on the size and dynamics of the workforce. The workforce is rapidly growing due to a rapidly growing population which is mainly a young population (about 51% are under the age of 25). The Saudi gross domestic product (GDP) fluctuates dramatically due to its close link to the price of oil.

Given this background which be easily deduced from a cursory reading of the news and simple searching, we now show how it can be further enriched through DBPedia through different queries.

Query 1: Find the value of GDP, the rate of inflation in the consumer price index (CPI) and the rate of population growth.

```
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?inflation_rate ?GDP_value ?population_rate

WHERE {

dbr:Economy_of_Saudi_Arabia dbp:inflation ?inflation_rate.
dbr:Economy_of_Saudi_Arabia dbp:gdp ?GDP_value.
dbr:Demographics_of_Saudi_Arabia dbp:growth ?population_rate

}
```

The above result (see Table IV) shows that the market value of all goods and services produced over a period of time is $1.68 trillion. The population is steadily increasing at a rate of 1.49 %; we would expect the GDP to increase with this rate.

TABLE IV. A partial result of query 1.

Inflation rate	GDP value	Population rate
3.0 %	1.679 trillion	1.49

Query 2: Find the population rate and total imports/exports of products and expenses.

```
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?Imports_value ?Imports_products ?Exports_value ?Exports_products ?Expenses ?Revenues ?population_rate

WHERE {
dbr:Economy_of_Saudi_Arabia dbp:imports ?Imports_value;
               dbp:importGoods ?Imports_products;
               dbp:exports ?Exports_value;
               dbp:exportGoods ?Exports_products;
               dbp:expenses ?Expenses;
               dbp:revenue ?Revenues.
dbr:Demographics_of_Saudi_Arabia dbp:growth ?population_rate
}
```

Results of the above query as shown in Table V indicate how the population rate affects expenses, revenue, imports/exports, and how imports/exports of products affect revenue. It also tells us that the Saudi market enjoys high positive net export in the international trade of petroleum products–particularly oil. Saudi imports include products such as machinery, foodstuffs, chemicals, motor vehicles and textiles, and the overall value of imports is $136.8 billion. The data indicates moreover that the population rate affects import values, imported products and expenses such that if the population rate increases these aforementioned values would also increase. The high value of exports would affect the revenue value ($171.6 billion). Given its high expenses and increasing population rate, one could expect that it would be a good idea to increase exports.

Table V. A partial result of query 2.

Imports value	Exports value	Expenses	Revenues
136.8 $	231.3 $	227.8 $	171.6 $

One purpose of querying the LOGD from DBpedia is to show the value of LOD and to study how the data that was already available in Wikipedia (in plain English language) could be queried when represented as RDF data. Another purpose is to show how the result of SPARQL queries can be helpful for defining future procedures and policies. Our study has indicates that if would be valuable to develop a framework to produce LOD from different sources (as described the LOSGD Framework) in Section IV.

B. Queries generated from multiple datasets

This section shows the usefulness of querying the EU LOD SPARQL endpoint and analyzing the results of these queries. We will also analyze the potential use of such linked data in developing and creating new and innovative applications, and link data from different sources.

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX odp: <http://data.europa.eu/euodp/ontologies/ec-odp#>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX bud: <http://data.europa.eu/bud#>
SELECT   ?TiAlias CONCAT(?Title_Heading," - ", ?Year) AS ?Head  ?Amount_a
FROM <http://data.europa.eu/euodp/resource/4rx/ABB2018>
WHERE { {
SELECT
str(?TiAlias) AS ?TiAlias
str(?TiHeading) AS ?Title_Heading
sum(xsd:decimal(?MV_Value)) AS ?Amount_a
str(?A_Year) AS ?Year
WHERE {{
    ?Title a bud:Title.
    ?Title bud:alias ?TiAlias.
    ?Title bud:heading ?TiHeading.
    FILTER (lang (?TiHeading) = 'en').
    ?Title bud:hasChapter ?Chapter.
    ?Chapter bud:hasArticle ?Article.
    ?Article bud:hasAmount ?Amount.
    ?Amount bud:year ?A_Year.
    #FILTER(str(?A_Year)="2018" OR str(?A_Year)="2017")
    ?Amount bud:hasPoliticalCategory ?A_CatPol.
    ?Amount a ?A_Type.
    ?Amount bud:figure ?MonetaryValue.
    ?MonetaryValue bud:value ?MV_Value.
    FILTER(REPLACE(xsd:string(?A_Type), '.*[/#]', '')="Commitment" OR REPLACE(xsd:string(?A_Type), '.*[/#]', '')="NonDifferenciated"))
UNION {
    ?Title a bud:Title.
    ?Title bud:alias ?TiAlias.
    ?Title bud:heading ?TiHeading.
    FILTER (lang (?TiHeading) = 'en').
    ?Title bud:hasChapter ?Chapter.
    ?Chapter bud:hasArticle ?Article.
    ?Article bud:hasItem ?Item.
    ?Item bud:hasAmount ?Amount.
    ?Amount bud:year ?A_Year.
    #FILTER(str(?A_Year)="2018" OR str(?A_Year)="2017")
    ?Amount a ?A_Type.
    ?Amount bud:figure ?MonetaryValue.
    ?MonetaryValue bud:value ?MV_Value.
    FILTER(REPLACE(xsd:string(?A_Type), '.*[/#]', '')="Commitment" OR REPLACE(xsd:string(?A_Type), '.*[/#]', '')="NonDifferenciated")}}}
GROUP BY(?TiAlias)
#HAVING(?Amount_a >100000000)
ORDER BY DESC (?Amount_a)   LIMIT 10
```

Query 3: Find the total amount earmarked per title and how does it change over the years.

This query retrieves the total amounts for each title over the years. The results show how the amount changes over the years (see Table VI), and that Agriculture and rural development have the highest total amount over years followed by Regional and urban policy.

Table VI. A partial result of query 3.

Head	Total amount
Agriculture and rural development-2016	63449875939.42
Agriculture and rural development-2018	58159838271
Regional and urban policy-2018	39812082371

Query 4: Find the total amount budgeted under each political category in 2018

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX odp: <http://data.europa.eu/euodp/ontologies/ec-odp#>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX bud: <http://data.europa.eu/bud#>

SELECT ' Amounts spent on ' AS ?Total_of
REPLACE(REPLACE(xsd:string(?Catpol), '.*[/#]', ''),' ','.') AS ?Political_Category
sum(xsd:decimal(?MV_Value)) AS ?Amount
FROM <http://data.europa.eu/euodp/resource/4rx/ABB2018>
WHERE {{
{ ?Article a bud:Article.
  ?Article bud:hasAmount ?Amount.
  ?Amount bud:year ?A_Year.
  FILTER(str(?A_Year)="2018")
  ?Amount a ?A_Type.
  FILTER(CONTAINS(str(?A_Type),"Commitment") OR CONTAINS(str(?A_Type),"NonDifferenciated")).
  ?Amount bud:figure ?MonetaryValue.
  ?MonetaryValue bud:value ?MV_Value.
  FILTER(str(?MV_Value)!='p.m.').
  ?Amount bud:hasPoliticalCategory ?Catpol.}

UNION
{
  ?Item a bud:Item.
  ?Item bud:hasAmount ?Amount.
  ?Amount bud:year ?A_Year.
  FILTER(str(?A_Year)="2018").
  ?Amount a ?A_Type.
  FILTER(CONTAINS(str(?A_Type),"Commitment") OR CONTAINS(str(?A_Type),"NonDifferenciated")).
  ?Amount bud:figure ?MonetaryValue.
  ?MonetaryValue bud:value ?MV_Value.
  FILTER(str(?MV_Value)!='p.m.').
  ?Amount bud:hasPoliticalCategory ?Catpol.}
}}ORDER BY DESC (?Amount)
LIMIT 10
```

This query retrieves the total amount budgeted for each political category in 2018. The result shows the budget for each political category in 2018 (see Table VII). The result indicates that the political category MFF2018_2_0_10 has the highest total amount in 2018 while the political category MFF2018_5_1_2 has the lowest total amount.

Table VII. A partial result of query 4.

Political category	Amounts
MFF2018_2_0_10	43234516899
MFF2018_2_0_11	27012257827
MFF2018_2_0_2	2366636521

We issued some queries related to governance to DBpedia and other related datasets. These results can be used to perform different types of analysis and develop applications. In general, when data from different data sources are linked (as the above shows) we can query the entire set from a single point using the SPARQL protocol. This is not possible without Linked Open Data. The result

of such queries can also be used for purposes emanating from different domains of knowledge.

VI. CONCLUSION AND FUTURE WORK

Open data initiatives are an important part of the contemporary sociopolitical vision of Saudi Arabia. They are a strong means of realizing aspirations of transparency, accountability, trust, and confidence sought of the Saudi governing and other public bodies in this time of unprecedented change and transition in multiple arenas. The effective development of the relevant technical components to ensure the success of these initiatives is therefore of great importance. The current situation where the provided open data is in multiple formats, locations and is of insufficient quality–which prevents effective computational analysis–needs to be changed as it is preventing the data from being utilized for policy making and hence fulfilling the required values. Our work addressed some of these limitations. We presented our Linked Open Saudi Government Open Data Framework (LOSGDF) for producing LOD from different data sources such as CSV files, Excel sheets, online portals, and structured documents and linking them to other existing open datasets, and facilitating intelligent querying through SPARQL.

This paper investigated the feasibility of current Saudi government open data in this regard and assessed existing approaches to improve the effectiveness of open government data through transforming it into linked-open data by connecting disparate sources of structured data therein. It proposed a framework for automating the linking sub-process of existing approaches and organizing the data to be queried through SPARQL and evaluated the potential benefit of this proposal by discussing some case studies demonstrating the kinds of analyses, insights, and policy decision such a framework could enable, which would be difficult without such a framework.

The immediate next task in this project would be to continue to enhance the implementation of this framework in different government sectors and investigate its ability to semantically enrich Saudi government data and produce linked open data from different sources.

REFERENCES

[1] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data: The story so far," in *Semantic services, interoperability and web applications: emerging concepts*, ed: IGI Global, 2011, pp. 205-227.

[2] A. Sáez Martín, A. H. D. Rosario, and M. D. C. C. Pérez, "An international analysis of the quality of open government data portals," *Social science computer review,* vol. 34, pp. 298-311, 2016.

[3] U. A. Algemili, "Outstanding Challenges in Recent Open Government Data Initiatives," *International Journal of e-Education, e-Business, e-Management and e-Learning,* vol. 6, p. 91, 2016.

[4] Ojo, Adegboyega K., Edward Curry and Fatemeh Ahmadi Zeleti. "A Tale of Open Data Innovations in Five Smart Cities." *2015 48th Hawaii International Conference on System Sciences* (2015): 2326-2335.

[5] Saudi Vision 2030. (2016). Government of Saudi Arabia.

[6] A. Zuiderwijk, R. Shinde, and M. Janssen, "Investigating the attainment of open government data objectives: Is there a mismatch between objectives and results?," *International Review of Administrative Sciences,* p. 0020852317739115, 2018.

[7] The Uncertain Relationship Between Open Data and Accountability: A Response to Yu and Robinson's *The New Ambiguity of "Open Government"* – Tiago Peixoto – UCLA Law Review Discourse 2013

[8] Tolbert, Caroline J., and Karen Mossberger. "The Effects of E-Government on Trust and Confidence in Government." Public Administration Review 66, no. 3 (2006): 354-69.

[9] Measuring the Data Openness for the Open Data in Saudi Arabia e-Government – A Case Study – AlRushaid, M and Saudagar, A. K. J. International Journal of Advanced Computer Science and Applications, Vol. 7, No. 12, 2016

[10] Motivations for open data adoption: An institutional theory perspective. Mohammed Saleh Altayar, Government Information Quarterly, Volume 35, Issue 4, October 2018, Pages 633-643

[11] Stuti Saxena, (2018) "National open data frames across Japan, The Netherlands and Saudi Arabia: role of culture", foresight, Vol. 20 Issue: 1, pp.123-134

[12] Crisis analytics: big data-driven crisis response. Qadir *et al. Journal of International Humanitarian Action* (2016) 1:12

[13] S. Mouzakitis, D. Papaspyros, M. Petychakis, S. Koussouris, A. Zafeiropoulos, E. Fotopoulou*, et al.*, "Challenges and opportunities in renovating public sector information by enabling linked data and analytics," *Information Systems Frontiers,* vol. 19, pp. 321-336, 2017.

[14] M. N. Spoiala, O. Rinciog, and V. Posea, "The semantic representation of open data regarding the Romanian companies," in *RoEduNet Conference: Networking in Education and Research, 2016 15th*, 2016, pp. 1-5.

[15] Kassen, M. (2018), Open data and its institutional ecosystems: A comparative cross-jurisdictional analysis of open data platforms. Can Public Admin, 61: 109-129

[16] C. Alexopoulos, E. Loukis, S. Mouzakitis, M. Petychakis, and Y. Charalabidis, "Analysing the characteristics of open government data sources in Greece," *Journal of the Knowledge Economy,* vol. 9, pp. 721-753, 2018.

[17] H. S. Al-Khalifa, "A Lightweight Approach to Semantify Saudi Open Government Data," in *Network-Based Information Systems (NBiS), 2013 16th International Conference on*, 2013, pp. 594-596.

[18] P. Fragkou, N. Kritikos, and E. Galiotou, "Querying Greek Governmental Site using SPARQL," in *Proceedings of the 20th Pan-Hellenic Conference on Informatics*, 2016, p. 80.

[19] E. Galiotou and P. Fragkou, "Applying linked data technologies to Greek open government data: a case study," *Procedia-social and behavioral sciences,* vol. 73, pp. 479-486, 2013.

[20] M. Vafopoulos, M. Meimaris, J. M. Á. Rodríguez, I. Xidias, M. Klonaras, and G. Vafeiadis, "Insights in global public spending," in *Proceedings of the 9th International Conference on Semantic Systems*, 2013, pp. 135-139.

[21] D. Misra and A. Mishra, "Societal and economical impact on citizens through innovations using open government data: Indian initiative on open government data," in *Handbook of Research on Cultural and Economic Impacts of the Information Society*, ed: IGI Global, 2015, pp. 147-178.

[22] 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 *Data and Software Engineering (ICODSE), 2014 International Conference on*, 2014, pp. 1-6.

[23] G. Jaglan and S. K. Malik, "LOD: Linking and Querying Shared Data on Web," in *2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2018, pp. 568-573.

[24] L. Boonlamp, "A linked data approach to planning collaboration amongst local governments in Thailand," in *Information Technology (INCIT), 2017 2nd International Conference on*, 2017, pp. 1-5.

[25] P. C. N. Azevedo, G. S. Bastos, and F. S. Parreiras, "A linked open data approach for visualizing flood information: A case study of the Rio Doce Basin in Brazil," in *Geographical Information Systems Theory, Applications and Management (GISTAM), 2015 1st International Conference on*, 2015, pp. 1-6.

[26] 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*, ed: Springer, 2007, pp. 722-735.

[27] J. McCrae, "The Linked Open Data Cloud", *Lod-cloud.net*, 2019. [Online]. Available: https://lod-cloud.net/.

Authors' Profiles

Afnan M. Alsukhayri received her Bachelor's degree in Information Technology from King Abdulaziz University, Jeddah, Saudi Arabia in 2013. Currently, she is studying to receive her master's degree in information system at the Faculty of Computing and Information Technology at King Abdulaziz University, Jeddah, Saudi Arabia. Her research interests are in the areas of Linked Open Data, Semantic Web and Social Network Analysis.

Muhammad Ahtisham Aslam received the Master degree in computer science from Hamdard University, Lahore, Pakistan, in 2002 and a Ph.D. degree from University of Leipzig, Leipzig, Germany in 2007. From 2003 to 2007, he had been working as Research Assistant and Ph.D. student at University of Leipzig, Leipzig, Germany. His research interest includes semantic web, linked open data, data science, and social network. Muhammad A. Aslam is currently working as Assistant Professor at Department of Information Systems, King Abdulaziz University, Jeddah, Saudi Arabia. He has been working as Senior Staff Researcher at Artificial Intelligence Lab, Knowledge Technology Cluster, Malaysian Institute of Microelectronic Systems (MIMOS), Kuala Lumpur, Malaysia. He also has been working as Assistant Professor at COMSATS Institute of Information Technology, Lahore, Pakistan.

Sachi Arafat is an Assistant Professor at the Department of Information Systems, King Abdulaziz University, Jeddah, Saudi Arabia. He was previously a Royal Academy of Engineering Postdoctoral Fellow at the School of Computer Science at the University of Glasgow, where he also completed his undergraduate degree and Ph.D. His work lay in the intersection of Information Science, Philosophy of Technology, Information Retrieval and Data Science. The most representative of this intersection is his book on 'Search Foundations' published in 2019 by the MIT Press.

Naif Radi Aljohani received his Bachelor's degree in Computer Education from King Abdulaziz University in 2005, the Master degree in Computer Networks from La Trobe University, Australia, in 2009 and a Ph.D. degree in Computer Science from the University of Southampton, UK, in 2014. Currently, Mr. Naif is working as Assistant Professor at the Faculty of Computing and Information Technology at King Abdulaziz University, Jeddah, Saudi Arabia. His research interests are in the areas of mobile and ubiquitous computing, mobile and ubiquitous learning, learning and knowledge analytics, semantic web, Web Science, technology-enhanced learning and human-computer interaction.