

A SoA for Open Government Data: A Case Study of COVID Impacts on SMEs

Ali Albinali¹, Russell Lock¹ and Iain Phillips¹

¹Loughborough University, Loughborough LE11 3TU, Leicestershire, UK

Abstract

The efficient utilization of Open Government Data (OGD) is one of the current major challenges for Small and Medium Enterprises (SMEs). OGD helps SMEs to find new business opportunities, offer high quality services and generate economic value. Current OGD platforms address issues such as data classifications and synchronization. Despite the extensive efforts to develop OGD platforms, there are still limitations. Existing platforms do not provide the ability for SME users to run complex queries which are based on data analytics techniques and algorithms. Also, they do not provide a smooth integration of data from different data sources. This paper introduces a Service-Oriented Architecture called the Data Analytics Framework (DAF) to design OGD platforms that provide functionality through provision of these services. The proposed framework is evaluated through a real life case study of COVID-19 impacts on SMEs, with specific reference to the use of sentiment analysis as an example data analysis technique applied to OGD.

Keywords

Service-Oriented Architecture (SOA), measurement, Open Government Data (OGD), Small and Medium Enterprises (SMEs), Data Analytics, Data Analytics Framework (DAF)

1. Introduction


Open data are already contributing to the economic growth of countries around the world [1]. They also support the creation and strengthening of new markets, organizations, and jobs [2]. Government plays an important role in the creation of value from open data, not only at the publication stage but also after deployment and when used during analysis by SMEs. Organizations can create value from open data in various industries [3]. Organizations use OGD to improve their performance and help in decision making. This also generates new products or services that generate value for the clients of these companies [4]. SMEs have a significant role in developing the economics of countries. Therefore, governments have attempted to develop OGD portals to provide new capabilities for SMEs to utilize [3]. However, governments still face several issues and limitations in developing OGD platforms, such as enabling various types of data analytics techniques and integration of data from multiple data sources.


SOA represents a significant breakthrough in the evolution of application development and integration. Service orientation splits problems into entity and related smaller parts of logic or service [5]. SOA provides efficient utilization of OGD through the deployment of multiple data analytics services.

EGOV-CeDEM-ePart 2022, September 06–08, 2022, Linköping University, Sweden (Hybrid)

✉ A.Albinali@lboro.ac.uk (A. Albinali); R.Lock@lboro.ac.uk (R. Lock); I.W.Phillips@lboro.ac.uk (I. Phillips)

🆔 - (A. Albinali); 0000-0002-4543-3753 (R. Lock); 0000-0001-5013-5876 (I. Phillips)

 © 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

The main objective of this paper is to introduce a SOA for OGD platforms. The OGD platform called Data Analytics Framework (DAF) provides a set of several data analytics which supported by a SOA architecture. We present a design for the overall OGD platform, but are only implementing a subset at this time, relating to Sentiment Analysis (SA) to prove the concept. The OGD platform is then applied to a case study that analyzes COVID-19 tweet data and their impact on SMEs in Qatar. The structure of the paper is as follows: We discuss the related work of OGD in section 2. Following this, the research methodology and the used data collection methods are briefly introduced in section 3. We present a suggested SOA architecture in section 4. We demonstrate the efficacy of the DAF using a case study to determine how COVID-19 affected SMEs in section 5. Then, we explore the implementation of SA as a set of services in section 6. We discuss the merits of our SOA design and the findings of the case study in section 7. Finally, we conclude our approach and outline directions for future research in section 8.

2. Related Work

OGD refers to the subset of open data that is government-related data open to the public [6, 7]. Government data contains diverse datasets such as finance, population, geographical, public, transportation, traffic and education. Several countries have already demonstrated their commitment to OGD by joining the Open Government Partnership [8]. The objective of this research is to develop a platform that answers complex queries from SMEs concerning OGD. Moreover, the platform should allow combining OGD with social media and other third party supplied data. We have examined many OGD initiatives from around the world selecting United States (data.gov), United Kingdom (data.gov.uk), Oman (data.gov.om), Qatar (data.gov.qa), India (data.gov.in) and Australia (data.gov.au) OGD initiatives due to number of reasons including *Geographic Locations* to cover all continents, *Maturity of the OGD Initiative* the degree of maturity and completeness of these OGD initiatives, and *Diversity of Selection* the different cultures from these various OGD initiatives.

Several studies have discussed the requirements for evaluating the OGD initiatives [9, 10, 11, 12]. The aspects are summarized and classified into four categories *Data, Portal, External factors and Public engagement* [9]. Moreover, other criteria are added such as *the context of the open data, and the perspectives*. The context of open data is classified into *Government, Public, Mixed, or General or Not defined* [10]. Welle Donker and van Loenen [11] introduced six assessment models *Open Data Benchmark, Scoreboard, Global Open Data Index, Tagging Framework, Maturity Framework and Open Data Barometer* [12]. We developed an evaluation approach to consolidate, classify and score the different aspects for each OGD Initiative. The evaluation approach is based on *criteria, criteria category or classification, aspect, value range, and Rank*. The criteria are categorized into technical (e.g. OGD Platform Installation, Configuration and Accessibility, OGD Platform Data Formats, OGD Platform Meta-data, and OGD Platform Data Analytics) and organizational (e.g. OGD Policies, OGD Lifecycle, Stakeholder Participation and Collaboration, and OGD Maturity). Each criteria has a set of aspects which has an associated range of values that we used to rank across the different OGD initiatives. After, we explored the different functionalities of OGD platforms, we compared and scored the different criteria and their aspects. We found that many current OGD initiatives are still in the early stages for

support of data analytics components. The existing OGD initiatives are not combining a data platform with an analytic platform. Also, they lack the ability to accept data from SMEs and unstructured data from social networks such as Twitter or Facebook.

3. Research Methodology

The Qatari Government needs to investigate the role of SMEs in Qatar to utilize and spread the use of OGD. Also, to explore the potential features of OGD platforms that existing OGD initiatives do not provide. Therefore, we made use of a mixture of quantitative and qualitative data collections such as *Survey*, *Interviews (Focus Group interviews)*, and *Experiment*. Such mixed data collection methods provide several merits: (results can be generalized to a bigger population, they are easier to analyze because the data are represented in a numerical form and the analysis can be displayed graphically) as quantitative methods and (analysis tends to be detailed in description, generates and test hypotheses and collected data using less structured research tools) for qualitative methods [13, 14]. We designed the research methodology to consist of five phases: *Data Collection - Survey*, *Data Collection - Interview and Focus Group Workshop*, *Apply OGD Maturity Model*, *Develop OGD Framework* and *Evaluate OGD Framework*.

Phase 1 - Data Collection - Survey: in this phase we developed two surveys for Open OGD: one for citizens and residents called the *Awareness Survey* which received 422 responses from 500 consumers, equating to a return rate of 84%. The other survey for investors and SMEs which was titled the *SMEs and Investors Survey* received 101 responses from list of 125 emails to SMEs requested from Ministry of Interior (MoI) State of Qatar services, equating to a return rate of 81%. The main finding from both surveys was that existing OGD platforms needs a solution for *integrating several organization's data to perform complex analysis scenarios or develop a required application*. The output of both surveys was a Market Need Report that was considered as an input for phase 2. **Phase 2 - Data Collection - Interview and Focus Group Workshop:** this phase started by investigating the Market Need Report from phase 1 through both interviews with organization's stakeholders such as managers and IT directors and focus group workshops with organization's stakeholders such as IT and Business Staff in these organizations. The key finding from both interviews and focus groups was that existing OGD platforms need to enable different types of data analytics (i.e. descriptive, diagnostic, predictive, and prescriptive). The OGD platform should answer the following queries using the suitable analytics; e.g. what are the impacts of COVID on SMEs in Qatar in a specific period on Twitter/Facebook, predict the price of apartment/house in a specific zone from labeled dataset, and classify if the customer will end a relationship with business/organization or not from labeled dataset. **Phase 3 - Apply OGD Maturity Model:** this phase assessed the maturity of the organization towards the application of OGD using a customized OGD Maturity Template as Open Data Maturity Model developed by Open Data Institute (ODI) [15]. The output of this phase are a set of requirements that each organization should achieve in order to reach a specific level of OGD Maturity levels. One of the significant requirements of OGD to be matured is the utilization of OGD using different analytics techniques. Detailed discussion of Phase 1 to Phase 3 are unfortunately out of scope due to paper length restrictions. **Phase 4 - Develop OGD Framework:** this phase collated the output from several requirements from phase 3. We then

developed a conceptual OGD DAF that satisfied these requirements. Moreover, we applied and implemented this framework for a selected organization(s) to validate the satisfaction of the requirements. SOA will enable us to provide a solution for these issues as presented in next section 4. **Phase 5 - Evaluate OGD Framework:** this phase evaluates our OGD framework and conclude its strengths and weaknesses.

4. SOA Architecture for OGD Platform

The findings listed in section 2 represent research which included the smooth integration of data from various data sources (i.e. data from SMEs and unstructured data from social networks), and around enabling several data analytics that answer SME users complex queries. There is a need for suitable design to satisfy these issues in existing OGD platforms. SOA is a useful concept in this context that is extensible and allows subject specific services to be interchanged as necessary. Our suggested SOA architecture, DAF, is used to support SMEs in their data-driven decision making for their business. Figure 1 represents a high-level architecture of the proposed DAF and its components. DAF analytics services consist of four main layers, *Dashboard, Query Editor, Schema and Data*. A brief description of these layers from bottom-up is presented as follows:

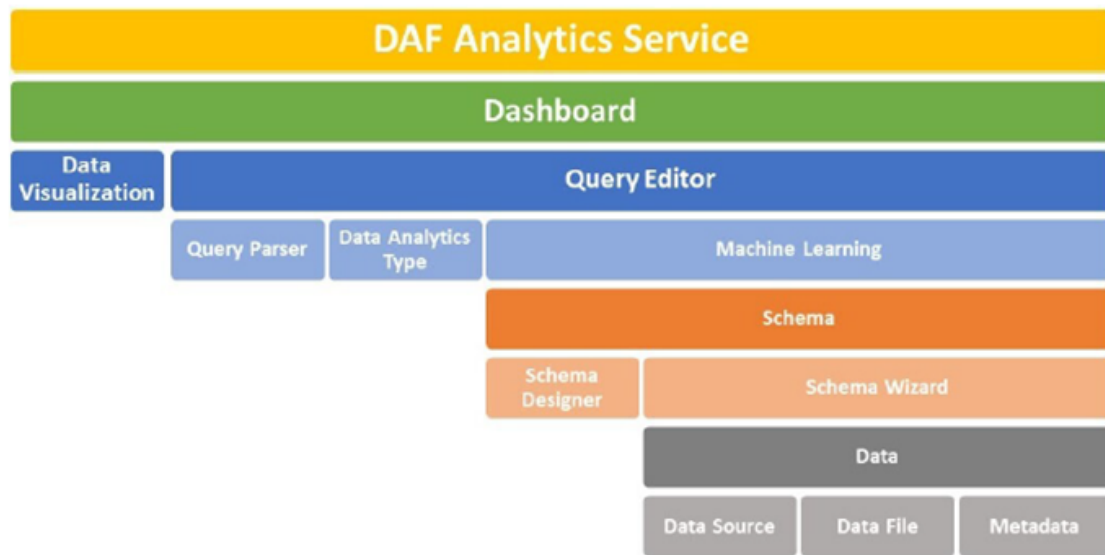


Figure 1: ODP Data Analytics Framework High-Level Architecture

1. **Data Services:** to extract and load data from a Data Source or a Data File or both. It consists of three components Data Source, Data File and Metadata [9]
 - **Data Source:** is a connection to database, or a software-as-service API.
 - **Data File:** is a structured data file such as Comma Separated Value file (.csv), Microsoft Excel file (.xls, .xlsx), or custom delimited file (.txt).

- **Metadata:** is the data associated with data sources or data files which require a classifier to enable the integration of the different organization's data.
2. **Schema Services:** defines structured, physical data as tables. It includes join relationships to other tables and views. It consists of two components Schema Wizard and Schema Designer.
 - **Schema Wizard:** to quickly detect the relationships between tables using the existing data sources to define tables.
 - **Schema Designer:** to manually load the tables and define the relationships between them.
 3. **Query Editor Services:** enables the SME user to write a query in English language. This enables people with minimal technical expertise to use the OGD platform. It consists of three components Query Parser, Data Analytics Type and Machine Learning (ML).
 - **Query Parser:** is the component responsible for translating the query of the user into an understandable query language such as Standard Query Language (SQL) and No-SQL for structured and unstructured data respectively.
 - **Data Analytics Type:** is the component responsible for identifying the analysis type required by the intended query.
 - **Machine Learning:** is the component responsible for selecting the most suitable ML algorithm to perform on the data retrieved by the query. DAF implements different types of ML such as supervised learning (e.g. classification and regression), unsupervised learning (e.g. clustering and dimensionality reduction), and reinforcement learning (e.g. real-time decisions, robot navigation, etc).
 4. **Dashboard Services:** enables SME user to use either Query Editor or/and Data Visualization
 - **Query Editor:** is the significant interface for SME users to write and perform their queries.
 - **Data Visualization:** is the component responsible for presenting the results to the SME users after applying the ML algorithm on the data retrieved by the query.

In the next section we present the details of our evaluative case study, which implements a subset of the framework discussed in this section.

5. Evaluation Case Study: Use of Sentiment Analysis

The case study scenario explores the impact of COVID-19 on SMEs through Twitter. In this scenario, we follow the journey of the SME user as in Figure 2. Firstly, the SME user writes a query in English language through the Query Editor Service. For example, the query could be "What are the impacts of COVID-19 on SMEs in Qatar in the period between 1st of January 2022 and 15th of February 2022 of on Twitter?". Secondly, a customized Natural Language Processing (NLP) algorithm as a service is utilized to extract the significant keywords which are mapped to both classification model and data source. For the classification model service, the

keywords such as analyze, impact, COVID-19, SMEs Qatar are important. For the data source, the keywords such as COVID-19, SME Qatar, Twitter, Period or dates. Thirdly, DAF apply the desired analytics or/and ML algorithms on the mapped data source(s). Finally, the outcomes of the query are saved into the Dashboard. In the final step, we need to extract data from Twitter as unstructured data. Moreover, Sentiment Analysis (SA) is the classification model which are suitable to answer the query defined in this scenario based on the extracted significant keywords. SA is an ongoing field of research to classify any text based on its polarity using text mining and NLP methods. NLP is a computational linguistic field concerned in understanding human languages.

The application areas of NLP involve several topics such as classification and clustering of documents, extraction of useful information (e.g. named entities), translation of text between and among languages, summarization of written works, automatic answering of questions by inferring answers [16]. SA is one of the classification problems that has gaining large interest with the increase of social media sites. People express their opinions and reviews on social media about a product, new campaign, an event, etc. Analyzing this huge amount of unstructured data of these social media provides useful information for any business or social considerations.

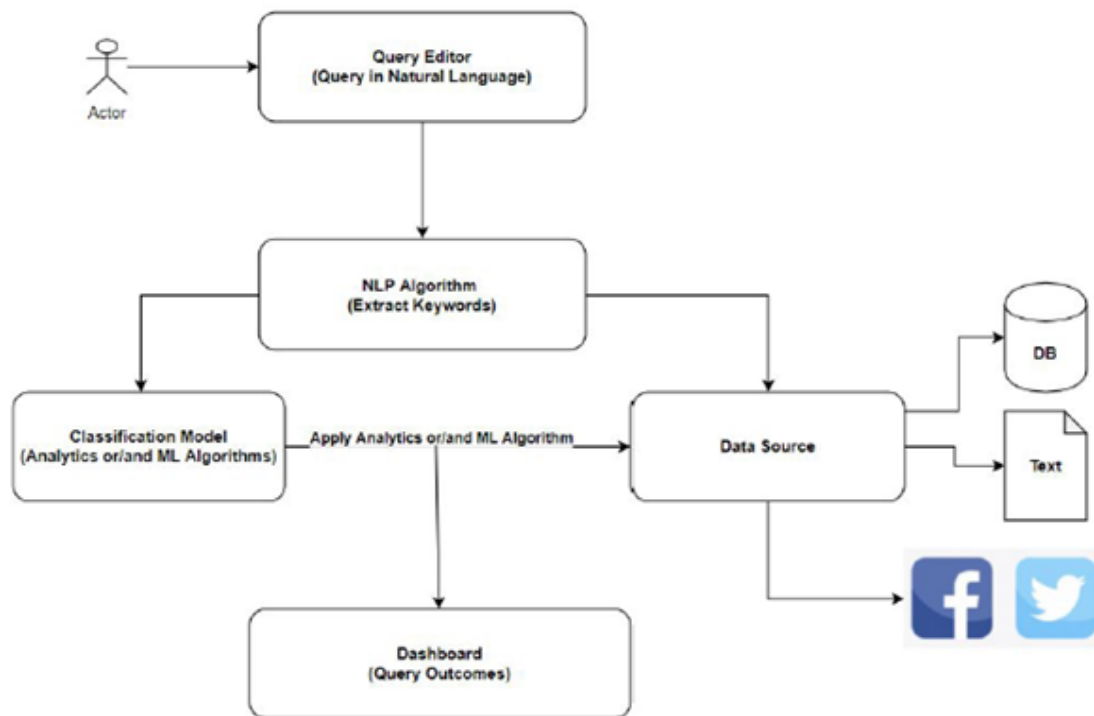


Figure 2: Detailed ODP Data Analytics Framework

Services Architecture is a detailed specifications for the tweet analysis services. Figure 3 shows the architecture of services to be provided for the SA and the interaction between them using UML Deployment Diagram. SA provides the following services: *Collection of tweet data*, *Pre-processing of the raw tweet to clean up text* and *Classification of tweet as positive or negative*.

Tweet Collector Service needs to specify the following: token to access API or rule ID for stop, Method (stream, stop), query for search, fields to be retrieved from the tweet and duration. Data is collected in json format and stored in a file with the rule ID. The **Pre-processor Service** takes a raw txt file collected from tweet API, and convert it into csv file with the given fields and clean the data based on a list of pre-processing methods. **SA Classifier Service** takes a csv file or text as input and extract features based on the specified algorithms and provide a csv file contains the polarity (positive, negative) or the score of the text.

Datasets SA models require large, specialized datasets to learn effectively. Datasets are available on a variety of topics (movies, tweets, hotels, books, etc.). Among popular datasets used for English SA; IMDB dataset is a Large Movie Review dataset [17]. Stanford Sentiment Treebank dataset contains user sentiment from Rotten Tomatoes, a movie review website [18]. The Sentiment140 dataset was collected using the Twitter API [19]. Yelp dataset with 4 million+ reviews [20]. Multiple datasets for Arabic SA are also available, such as: Arabic Jordanian General [21]. Arabic Sentiment Tweets [22]. Arabic Sentiment Twitter dataset for LEVantine dialect [23]. Hotel Arabic-Reviews Dataset collected from Booking.com [24]. A Large-SCaleArabic Book Reviews dataset [25]. SS2030: An Arabic Saudi tweets and is manually labelled [26].

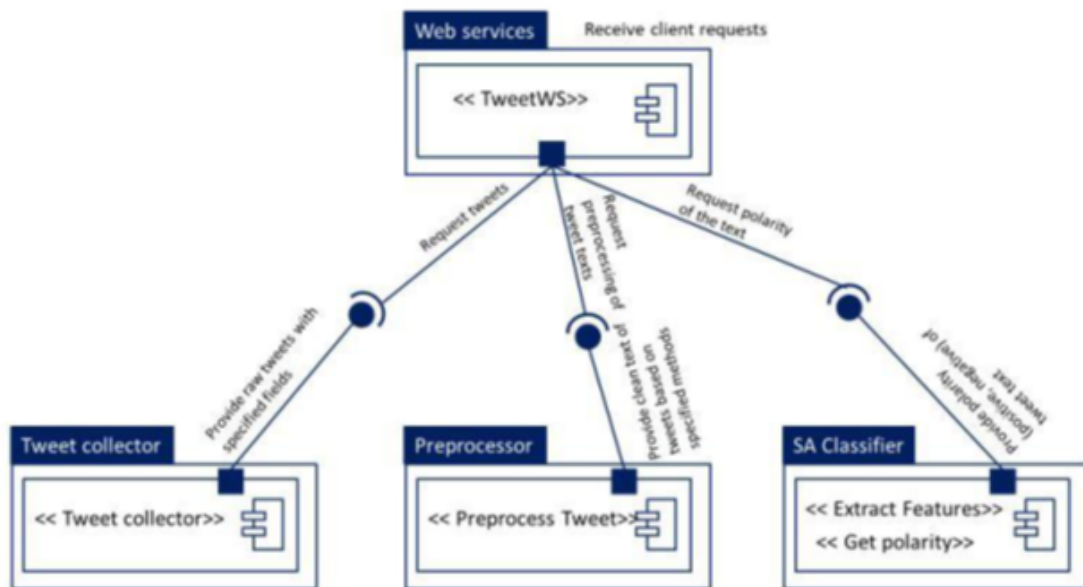


Figure 3: Sentiment Analysis Services Architecture

6. DAF Implementation: Tweet Analysis Services

The empirical material in this paper comprises one main survey document from open comparison and information on the websites This section introduce the sentiment analysis as one of several services provided by DAF. DAF implementation is based on SOA that could be reused for

different queries and analytics. The functionality required for processing is generic, and subject to the application of suitable rules can usefully be embodied as generic services within a SOA, lowering the boundaries to data analysis for SMEs. It answers the query of the case study in section 5. In text, several steps are performed to extract useful information. The first step is to collect data from social media about any specific brand, product or topic. The data collected are unstructured, which involves text pre-processing step to clean it. Among pre-processing, we distinguish several tasks like removing stop words, lowercasing, stemming, etc. depending on the use cases. After cleaning the data, it is necessary to convert it into number or vectors of numbers required by ML algorithms which called feature extraction. The last step is to apply the classification algorithm and to get the sentiment polarity. Both lexicon and ML-based approach have been proposed for SA.

Collection of Tweet Data Tweets are a specific kind of data carrying opinions on various topics, such as political parties, stocks, etc. The collection of twitter data can be done via the help of (Twitter API). It can be used to programmatically retrieve and analyze data, as well as engage with the conversation on Twitter. The newest Twitter API v2 supports additional features, metrics and access. By default, the data is collected in json format; it can be changed to any other formats for easy accessibility. **Text Pre-processing** it transforms the text into a form that is predictable and analyzable by ML algorithms. Some of the common text pre-processing/cleaning steps are: lower casing, removal of punctuation [!"] etc.], removal of stopwords defined by the nltk library, Stemming eliminating affixes from a word to obtain a word stem. Porter Stemmer is the most widely used technique because it is very fast (e.g. Working →Work), Lemmatization returns the base or dictionary form of a word, also known as the lemma (e.g. Better →Good), Tokenizing to turn the tweets into tokens. Tokens are words separated by spaces in a text, removal of frequent words, removal of rare words, removal of emojis, conversion of emoticons to words, spelling correction, etc [27]. In tweet data, additional pre-processing may be involved such as: *removal of hashtags, removal of mentions, and removal of specific words*. **Features Extraction** in ML algorithms, it is necessary to convert the set of texts into some vectors of numbers called features that can be fed into the model for processing. Depending upon the usage, features can be extracted using various techniques: Bag-of-Words (BoW), *Term Frequency-Inverse Document Frequency (TF-IDF)*, *Word embedding (word2vec, GloVe)* [27]. **Sentiment Classification Algorithms** to detect and extract emotions using ML, Lexicon-Based Approach and Hybrid Approach. Each tweet will scored and labeled either as *Positive, Negative, or Neutral* as an output of the SA algorithm.

7. Discussions and Findings

Due to the space restrictions of the paper we are unable to illustrate the connection between the results of the queries and the findings. This section summarizes the findings from applying SOA approach for OGD platforms and demonstrating the approach using the described case study. The results are divided into three findings as follows: Firstly, the impact of COVID-19 on SMEs in Qatar appears in several actions or decisions related to the business and processes of organizations. For example, human resource departments for many SMEs replaced the normal hiring process from outside Qatar to outsourcing hiring. Other Qatari SMEs changed

their business model to satisfy their customer needs. Therefore, these decisions need more analysis and proofs and check SMEs in other countries also. Moreover, privacy issues and data protection regulations may differ from country to another for using social media data and performing analytics. Secondly, SOA enables several advantages for such as *reliability, location independence, scalability, reusability, and easy maintenance* for the OGD platform. Small and independent services in the SOA enables testing and debugging the applications easily instead of massive code chunks which provides high applications reliability. Location independence enables changes to service locations over time without interrupting consumer experience on the system. Scalability enables services to run across multiple platforms, and programming languages. Reusability allows the accumulation of small, self-contained and loosely coupled functionality services. Easy maintenance of the application has become far easier without having to worry about other services. Finally, the SA of COVID-19 Tweets, and what are the impacts happened to SMEs. A fusion algorithm is implemented to combine the result of two SA classifiers [28, 29]. It improved the accuracy by one percent rather than using BERT or BiLSTM separately. However, the partial implementation of DAF needs to be extended to support further generically applicable data analytics techniques. Moreover, OGD from various sources requires smooth integration.

8. Conclusions and Future Work

This paper introduced DAF as SOA approach for OGD platforms. The DAF helps both government and SMEs in publishing and utilizing the OGD. The most significant finding behind the approach is the advantages of using SOA such as *reliability, reusability, scalability, etc.* Moreover, the application of our approach to a real world case study. Therefore, the characteristics of the case dominate in deciding the most suitable data analytics technique. In this paper, we designed a SOA for OGD that is based on different services such as data, schema, query editor, visualization and dashboard. Then, we validated the design of SOA through a case study. In future work, we need to implement additional services for data analytics, and other DAF services such as schema, data visualization and dashboard. We seek to find more data analytics techniques and apply the approach for more real world cases in different domains such as IT, Healthcare, Education. Also, we need to consider the feedback from SMEs and the responsible authorities of OGD. Finally, a proof-of-concept prototype for several data analytics techniques that validates the concept behind approach will be implemented.

References

- [1] A. Stott, Open Data for Economic Growth, Technical report,, The World Bank, 2014.
- [2] G. Michener, O. Ritter, Comparing resistance to open data performance measurement: public education in brazil and the uk, Public Administration 95 (2017) 4–21.
- [3] R. Fernandez, S. Ali, Sme contributions for diversification and stability in emerging economies-an empirical study of the sme segment in the qatar economy, Journal of Contemporary Issues in Business and Government 21 (2015) 23–45.

- [4] T. Jetzek, M. Avital, N. Jørn Andersen, Data-driven innovation through open government data, *Journal of theoretical and applied electronic commerce research* 9 (2014) 100–120.
- [5] N. Niknejad, A. Hussin, I. Amiri, Literature review of service-oriented architecture (SOA) adoption researches and the related significant factors, in: *The Impact of Service Oriented Architecture Adoption on Organizations*, Springer, 2019. URL: <https://doi.org/10.1007/978-3-030-12100-6>. doi:10.1007/978-3-030-12100-6.
- [6] J. Kučera, D. Chlapek, M. Nečaský, Open government data catalogs: Current approaches and quality perspective. Kó et al, in: eds. (Ed.), *EGO-VIS/EDEM 2013, Technology-Enabled Innovation for Democracy, Government and Governance*, LNCS, volume 8061, Springer, Berlin Heidelberg, 2013, p. 152–166.
- [7] A. Bachtiar, S. Suhardi, W. Muhamad, *International Conference on Information Technology Systems and Innovation (ICITSI (2020))* 329–334.
- [8] Open Government Partnership, 2011. URL: <https://www.opengovpartnership.org>.
- [9] J. Attard, F. Orlandi, S. Scerri, S. Auer, A systematic review of open government data initiatives, *Government Information Quarterly* 32 (2015) 399–418. URL: <https://doi.org/10.1016/j.giq.2015.07.006>. doi:10.1016/j.giq.2015.07.006.
- [10] M. Hossain, Y. Dwivedi, N. Rana, State-of-the-art in open data research: Insights from existing literature and a research agenda, *Journal of Organizational Computing and Electronic Commerce* 26 (2016) 14–40. URL: <https://doi.org/10.1080/10919392.2015.1124007>. doi:10.1080/10919392.2015.1124007.
- [11] F. Welle Donker, B. Loenen, How to assess the success of the open data ecosystem?, *International Journal of Digital Earth* 10 (2017) 284–306. URL: <https://doi.org/10.1080/17538947.2016.1224938>. doi:10.1080/17538947.2016.1224938.
- [12] H. Lindén, J. Stråle, An Evaluation of Platforms for Open Government Data, Technical report, KTH, School of Technology and Health (STH), Hanen, Sweden, 2014.
- [13] J. Creswell, D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 5th ed., SAGE Publications, Inc, 2018.
- [14] M. Saunders, P. Lewis, A. Thornhill, *Research Methods for Business Students*, 4th ed., Pearson Education Limited, 2007.
- [15] L. Dodds, A. Newman, A guide to the Open Data Maturity Model - Assessing your open data publishing and use, Technical report, Open Data Institute (ODI), UK, 2015.
- [16] D. Otter, J. Medina, J. Kalita, A survey of the usages of deep learning for natural language processing, *IEEE Transactions on Neural Networks And Learning Systems* (2019).
- [17] Large movie review dataset (lmrd), n.d. URL: <https://ai.stanford.edu/~amaas/data/sentiment/>.
- [18] NLP Stanford Sentiment Analysis, 2022-03-15. URL: <https://nlp.stanford.edu/sentiment/code.html>.
- [19] About twitter api, n.d. URL: <https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api>.
- [20] Yelp Dataset, 2022. URL: <https://www.yelp.com/dataset>.
- [21] AJGT Dataset, 2017. URL: [https://metatext.io/datasets/arabic-jordanian-general-tweets-\(ajgt\)](https://metatext.io/datasets/arabic-jordanian-general-tweets-(ajgt)).
- [22] M. Nabil, M. Aly, A. Amir, Astd: Arabic sentiment tweets dataset, in: *2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Lisbon, Portugal, 2015, pp.

2515–2519,.

- [23] ArSentD-LEV, 2022. URL: <https://paperswithcode.com/dataset/arsentd-lev>.
- [24] Hotel arabic reviews dataset (hard), n.d. URL: <https://github.com/elnapara/HARD-Arabic-Dataset>.
- [25] Labr: A large-scale arabic book reviews dataset, n.d. URL: <https://github.com/mohamedadaly/LABR>.
- [26] S. Alyami, S. Olatunji, Application of support vector machine for arabic sentiment classification using twitter-based dataset 19 (2020) 1–13. URL: <https://doi.org/10.1142/S0219649220400183>. doi:10.1142/S0219649220400183.
- [27] K. Kowsari, K. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, D. Brown, Text classification algorithms: A survey, *Information 10* (2019). URL: <https://doi.org/10.3390/info10040150>. doi:10.3390/info10040150.
- [28] Bert, 2022. URL: <https://github.com/google-research/bert>.
- [29] Understanding LSTM Networks, n.d. URL: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>.