

Data Citation Service for Wikipedia Articles

Arif Bramantoro¹, Ahmad A. Alzahrani²

¹Faculty of Computing and Information Technology Rabigh, King Abdulaziz University, Rabigh, Saudi Arabia

²Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract—The citation of big scientific data is crucial not only for scientific activity but also for the scientific discovery and dissemination within scientist network. The main objective of this research is to develop a service-oriented data citation system using data mining techniques for Middle East and North Africa scientists. A novel service oriented framework is proposed to prototype the development of the system that includes query formalization, service discovery, service composition design, service selection, search space, and service optimization. In this research, Wikipedia scientific-related articles are connected with more than 35 petabyte Pangaea datasets. The output of this work is a web service that takes Wikipedia article information as an input and provides the possible relevant datasets (if exist) related to the article. The evaluation of this research is based on a quantitative assessment performed to the quality of web service metrics, such as number of access and bandwidth utilization; which shows that the framework is robust enough to handle both big data access and its citation.

Keywords—Scientific dataset; web services; wikipedia; pangaea; big data

I. INTRODUCTION

Scientific data residing in datasets are usually considered as a shared resource, so that world-wide scientific community can access these datasets for their relevant purposes. Similar to journal papers and conference papers, which are cited in scientific papers, datasets are also significant candidates for citations in scientific papers/articles. Data citation is the reference or link from one piece of content to other contents in the form of data or document. There are many big data centers in which these kinds of datasets reside. Pangaea is one of data publishers for scientific data, mostly from earth and environmental science. It is open to any scientists or projects who are willing to archive and publish data. This online scientific resource repository gives an access to huge number of (mostly free) datasets for more than 35 PB in size. Wikipedia, on the other hand, is a gigantic collaborative encyclopedia consisting of huge number of articles, including a scientific category. The scientific-related articles of Wikipedia, just like any other scientific articles, have a particular section of citations. However, data citations section is not provided in Wikipedia. This is because to find the relationship between Wikipedia articles and scientific big data is not an easy task.

Data mining technique to find the connection between the two huge resources of data i.e. Wikipedia and Pangaea, is proposed in this project. The application of data mining on these big data produces the relationship between a Wikipedia article and a Pangaea dataset. There might be a debate on whether Wikipedia articles can be easily connected to Pangaea datasets or not, but it certainly requires additional layers of work. It means that there must be an application of some

intelligent services, so that Wikipedia article can be linked to Pangaea dataset.

It is believed that the research approach adopted in this study constitutes an appropriate and important way to reach international audience by the help of web service, particularly used by scientists in Middle East and North Africa region. One can input information of Wikipedia article and the web service outputs the relevant Pangaea dataset, if exist, together with some statistical measures on its performance. This service is considered to be helpful for anyone who wants to enhance Wikipedia article by editing it and adding data citation for relevant data sets. In addition, this service is helpful for scientists to collect Wikipedia articles related to their research interests and downloading the cited scientific data according to their research purposes.

The main objective of this research is to develop an automatic service-oriented data citation system, utilizing the model produced by data mining techniques. In detail, the objectives are as follows:

- 1) To develop a robust mechanism to connect scientific-related articles with scientific dataset.
- 2) To provide service-oriented framework for citing large-scale data.
- 3) To give attribution of the contributors of scientific data and written web documents.

II. RELATED WORKS

Scientific data citation has been attracting many researchers not only in the field of computer science [1]. There have been enormous efforts in dealing with big data in e-science [2]. However, some challenges remain exist [3]. One of the challenges is dealing with data too big to analyze, as highlighted in [4]. To date, two technological breakthroughs are available to overcome this issue, grid computing [5] and MapReduce [6]. Because MapReduce is the latest technology and proven to be robust in archiving scientific datasets, it is necessary to utilize MapReduce to store big data of scientific dataset, before analyze it.

An automatic data citation idea has been previously proposed as a poster in [7]. Although it is still in the conceptual idea and no implementation yet, the authors in [7] give us a rough idea that automatic data citation is very important. In this research, it is necessary to implement the same idea to Wikipedia scientific-related articles by utilizing a robust data mining technique, i.e. association rule discovery, applied find the relationships between text attributes provided by Wikipedia articles and data attributes provided by scientific dataset.

Another similar effort of dealing with scientific workflow has been proposed in [8]. A novel rule based approach model

and infrastructure is considered as the best approach for handling a huge explosion in scientific data. This is due to the internal requirement in processing the streaming of scientific data, especially through social media and sensor. This approach is suitable for real time and emergency data, but not necessarily required for data citation. Data citation is mostly processed through metadata without mining the big scientific data.

The effort of utilizing Wikipedia articles to understand a better science has been done by Mietchen and others [9]. The authors created a bot that is able to search automatically the multimedia files of medical domain and upload these files to Wikipedia media repository. There are approximately more than ten thousands files attached to hundreds Wikipedia articles by the bot. It works by exploiting XML tag values of the files. Neither semantic capability nor data mining technique has been exploited by the bot.

The first definition of data service is coined in [10] which aims at seeking the technological trends in simple, united and cloud-based data service. The authors defined data services as the offsprings inherited from stored procedure in relational database management system that any programmers and database administrators are able to code both SQL queries and programming control logic in one place in order to provide both query and function optimization capability for getting at the data, as normally alleviated in procedural programming languages compilation. This stored procedure is used as an analog for data service since only stored procedure are able to get an access to important and sensitive data to some cases. The authors also characterized data service as a service based on a proprietary model. The advantage of this definition is that there will be more enhanced perspective of the data as well as more data oriented. In addition, the architecture is extended to fit the characteristics of data citation service.

Perhaps the most similar work to this research has been proposed in [11], which provides application-oriented search. The work is exploiting the use of scientific metadata generated by scientific experiment. With the power of indexing, the new search is not only relying on keyword based matching, but also semantic web and its ontology to increase the accuracy of searching. In this research, it is argued that the use of data mining technique is more appropriate to analyze the big scientific data. This is due to the need of Wikipedia articles in relating the scientific data. Indexing scientific metadata is not the focus in this research, but the scope is to analyze the relationship between big scientific data and Wikipedia articles.

III. SCIENTIFIC DATASET

“As research on so many fronts is becoming increasingly dependent on computation, all science, it seems, is becoming computer science” announced by the New York Times in a 2001 famous article [1]. Recently, many organizations have accumulated data from various sources like Web, network sensors and constructed large-scale data. Some organizations publish their data to public to facilitate activities of other organizations. For example, WDS (World Data System)¹, science data in a wide range of domains has been registered. Another example is Pangaea², a system which provides a part of science data

on WDS related to the earth and environment, contains 0.6 million datasets and 35 PB data in total. These 0.6 million datasets can be referenced by their DOIs for direct access and citation.

Pangaea is chosen in this research, because it is considered as the biggest repository for scientific data, focusing on earth and environmental science. This enables us to contribute to the local scientific society since earth and environmental data are usually specific to the geographical area. Another advantage is that Pangaea provides a comprehensive set of metadata to make us easier to identify and analyze datasets. For example, one of datasets residing at Pangaea has following title: “Contents of rare earth elements and some rock-forming chemical elements in bottom sediments from some deeps of the Red Sea.”

Fig. 1 depicts what type of information is available for a particular dataset. The attributes of the dataset are also present in the webpage³ from which snapshot is taken but they are not shown here in order to save space.

Wikipedia is a collaborative encyclopedia consisting of millions of articles. There are scientific-related articles that can be expanded by connecting them to relevant dataset residing at Pangaea. This makes the Wikipedia articles more informative and useful, such as connecting the dataset discussed above with relevant Wikipedia article. Since the dataset discusses about Red Sea Minerals, it requires a validation whether Wikipedia has such information in any of its articles or not. Further search reveals that there exists an article on red sea in Wikipedia in which a section is there that discusses about the minerals found in the depth of Red sea as illustrated in Fig. 2.

Another section of the article which discusses about the minerals in Wikipedia article is illustrated in Fig. 3. Hence, the relation between the Wikipedia article and scientific dataset can be in the form of one to one, one to many or many to many, that requires the flexibility of the system to cite on the fly.

After manually checking the two resources, it can be inferred that such dataset from Pangaea can be cited in the above Wikipedia article. However, there is an urgent requirement to have an automatic citation for millions of Wikipedia articles related to the big data of scientific dataset based on heterogeneous, loosely coupled and platform independent application by utilizing web service technology [12].

IV. METHODOLOGY

To achieve the research objective, the following steps are envisioned.

- Development of Corpus.
In this step, full Wikipedia articles are fetched as dumps into the database. Moreover the full metadata for every datasets from Pangaea is also downloaded. This corpus is expected to lead us to venue of Big Data.
- Preprocessing of Data.
Since the articles are in the form of text, lot of text processing techniques are applied to make it suitable

¹<http://www.icsu-wds.org>

²<http://www.pangaea.de>

³<http://doi.pangaea.de/10.1594/PANGAEA.774361>



PANGAEA®
Data Publisher for Earth & Environmental Science

Always quote citation when using data!

Data Description

Show Map Google Earth

Citation: Butuzova, GY, Lyapunov, SM (1995): Contents of rare earth elements and some rock-forming chemical elements in bottom sediments from some deeps of the Red Sea. Geological Institute, Russian Academy of Sciences, Moscow, doi:10.1594/PANGAEA.774361,
In Supplement to: Butuzova, Galina Yu; Lyapunov, Sergey M (1995): Rare earth elements in hydrothermal-sedimentary deposits from the Red Sea. Translated from: *Litologiya i Poleznyye Iskopaemye*, 1995, 1, 16-30, *Lithology and Mineral Resources*, N 1, 13-26

Project(s): Archive of Ocean Data (ARCOD) a

Coverage: Median Latitude: 20.722289 * Median Longitude: 38.362489 * South-bound Latitude: 17.888300 * West-bound Longitude: 38.053333 * North-bound Latitude: 21.388300 * East-bound Longitude: 39.813300
Minimum DEPTH, sediment/rock: 1.1 m * Maximum DEPTH, sediment/rock: 10.6 m

Event(s): AK22-1996 a * Latitude: 21.283333 * Longitude: 38.053333 * Elevation: -2205.0 m * Location: Red Sea a * Device: Gravity corer (GC) a
PS-226 a * Latitude: 17.888300 * Longitude: 39.813300 * Elevation: -886.0 m * Location: Red Sea a * Campaign: PIKAR a * Device: Gravity corer (GC) a
PS-363 a * Latitude: 21.388300 * Longitude: 38.061700 * Elevation: -2085.0 m * Location: Red Sea, Atlantis II Deep a * Campaign: PIKAR a * Device: Gravity corer (GC) a



Fig. 1. Metadata of scientific dataset from Pangaea.

Red Sea

From Wikipedia, the free encyclopedia

Coordinates: 22°N 38°E

This article is about the body of water between Arabia and Africa. For other meanings, see Red Sea (disambiguation).

The **Red Sea**, or what is sometimes called the **Erythraean Sea**, is a **seawater inlet** of the **Indian Ocean**, lying between **Africa** and **Asia**. The connection to the ocean is in the south through the **Bab el Mandeb** strait and the **Gulf of Aden**. In the north, there is the **Sinai Peninsula**, the **Gulf of Aqaba**, and the **Gulf of Suez** (leading to the **Suez Canal**). The Red Sea is a **Global 200 ecoregion**. The sea is underlain by the **Red Sea Rift** which is part of the **Great Rift Valley**.

The Red Sea has a surface area of roughly 438,000 km² (169,100 mi²).^{[1][2]} It is about 2250 km (1398 mi) long and, at its widest point, 355 km (220.6 mi) wide. It has a maximum depth of 2211 m



Fig. 2. An example of Wikipedia article possibly corresponds to scientific dataset from Pangaea.

Geology

The Red Sea was formed by the **Arabian peninsula** being splitting from the **Horn of Africa** by movement of the **Red Sea Rift**. This split started in the **Eocene** and accelerated during the **Oligocene**. The sea is still widening, and it is considered that it will become an ocean in time (as proposed in the model of **John Tuzo Wilson**). In 1949, a deep water survey reported anomalously hot brines in the central portion of the Red Sea. Later work in the 1960s confirmed the presence of hot, 60 °C (140 °F), saline brines and associated metalliferous muds. The hot solutions were emanating from an active seafloor **rift**. The high salinity of the waters was not hospitable to living organisms.^[16] Sometimes during the **Tertiary** period, the **Bab el Mandeb** closed and the Red Sea evaporated to an empty hot dry salt-floored sink. Effects causing this would have been:

- A "race" between the Red Sea widening and **Perim Island** erupting filling the Bab el Mandeb with **lava**.
- The lowering of world **sea level** during the **Ice Ages** because of much water being locked up in the **ice caps**.

A number of volcanic islands rise from the center of the sea. Most are dormant. However, in 2007, **Jabal al-Tair Island** in the Bab el Mandeb strait erupted violently. An eruption among the nearby **Zubair islands** followed in 2011.^[19]

Mineral resources

In terms of mineral resources the major constituents of the Red Sea sediments are as follows:

- Biogenic constituents:
Nanofossils, **foraminifera**, **pteropods**, siliceous fossils
- Volcanogenic constituents:
Tuffites, **volcanic ash**, **montmorillonite**, **cristobalite**, **zeolites**
- Terrigenous constituents:



Fig. 3. Another part of Wikipedia article that discuss the scientific dataset in detail.

for data mining techniques. The Wikipedia articles require special preprocessing in order to produce output that is used as input for next step.

Since there is no training data, association rule mining algorithms as well as different similarity measurement techniques are applied to find the related articles.

- Application of different Data Mining Techniques.
- Web Service for World Wide Audience.

A web service service is dedicated to the users who are interested in finding the relevant scientific datasets for a particular Wikipedia article. It has the functionality of providing the Wikipedia article information. The output of service is the relevant and related Pangaea datasets.

To realize the expected results, the following mechanism is proposed.

- 1) A properly archived corpus of Wikipedia articles and Pangaea metadata.
To obtain the corpus, an open source MapReduce based Hadoop archive system is employed.
- 2) An ordered association and relationship information between Wikipedia scientific-related articles and Pangaea big datasets.
To obtain the information, a robust data mining of Association Rule techniques are employed.
- 3) Scientific Data Citation Service for Wikipedia scientific articles, Pangaea big datasets, and the relationship information that supports platform independent, autonomous, dynamic, reusable, heterogeneous, self-contained and loosely-coupled system.
To develop the service, XaaS (Everything as a Service) techniques, including DaaS (Data as a Service), TaaS (Text as a Service), CaaS (Citation as a Service), and SaaS (Software as a Service) are employed.

The mechanism of XaaS is implemented in service-oriented framework to ease the prototyping of the data citation system development as proposed in Fig. 4. The framework starts with query formalization and decomposition to ease the search for scientific dataset. The service discovery follows the step by providing more semantic capability during the searching process in multi-ontological environment [13]. The design of service composition improves a simple pattern of searching into fully automatic composition. The service selection follows the process until it combines Wikipedia and Pangaea datasets. The searching space is eventually improved from simple search by utilizing the previous approach of service atomization [14]. Finally, the simulation is conducted in several stages during service optimization of first algorithm.

With this research, it is expected that in long term, there will be an increase of productivity of the scientists in the Middle East and North Africa regions in term of re-do experiments, fast and easy access to Wikipedia scientific-related articles and the corresponding Pangaea scientific datasets. It is also expected there will be an improved management for information services in the Kingdom due to integral administration of data citation as an information asset. This can be done through the expected results of this project that provides useful relationship information between Wikipedia articles and scientific data, in order to be utilized by the scientists.

The data are analyzed to mine the relationship between two different objects of properly archived corpus of Wikipedia articles and Pangaea metadata. The scientists can utilize the corpus for their research. For example, Wikipedia scientific article corpus is used for a layman-friendly guidelines about the research topics and the Pangaea metadata corpus is used for re-do their experiments.

An ordered association and relationship information between Wikipedia scientific-related articles and Pangaea big datasets are provided to the scientist to choose the right Pangaea datasets for particular Wikipedia scientific-related articles. The association and relationship information is ordered based on the degree of the relatedness between Wikipedia scientific-related articles and Pangaea big datasets, so that the scientists have a variety of options to choose the appropriate datasets for particular Wikipedia articles.

Scientific Data Citation Service for Wikipedia scientific articles, Pangaea big datasets, and the relationship information web services are provided to make the information publicly available and seamlessly integrated to any legacy systems. The relationship service between Wikipedia scientific-related articles and Pangaea big datasets is utilized by scientists to cite the most related datasets for the articles. Archive service of Wikipedia scientific articles are provided to scientists to search and access the required articles based on the datasets owned or accesses by scientists. Archive service of Pangaea big datasets are provided to scientists to search and access the required datasets based on the Wikipedia scientific articles written or accesses by scientists.

V. DATA MINING AND ITS ARCHITECTURE

This research proposes to utilize data mining concept in this research, due to the nature of big data. Data mining is able to identify what data are passed between services. Since there are several services in this research and collaboration amongst the services in one composition, the identification of data is non-trivial. Data citation service also requires data mining to know what services are available as well as what results are generated for particular sets of input values. More specifically, data mining enables user to trace the process that led to the aggregation of services producing particular output.

Data mining processes are combined in the form of graph, which is implemented later as rule or workflow. The graph contains hybrid data and services. There is no focus on this data mining technique, since all available data and services are grouped together depends on the aim of the model. For example, the Fig. 5 shows data citation service utilizing three data mining services and seven datasets. Based on the objective of the workflow, the data mining technique determines that there should be two groups of service S_2 and S_3 , before composed with service S_1 . The group of service S_2 are related with dataset D_1 and D_3 , whereas the group of service S_3 are correlated with data set D_3 , D_2 and D_2' (a complimentary of dataset D_2).

Different objective has different composition. Fig. 6 illustrates that similar workflow may have different data mining tools. All services and datasets are in the same group, except for all datasets complimentary, D_0' and D_2' .

To implement the data mining tools in data citation service, a data citation service architecture is designed to accommodate the nature of scientific services and dataset as illustrated in Fig. 7. This data citation service architecture is based on the basic data service architecture proposed in [10]. The architecture extension includes an enhancement of the traditional service where the operations of the service, i.e. inputs and outputs, are semantically unplugged to the clients; which has no data

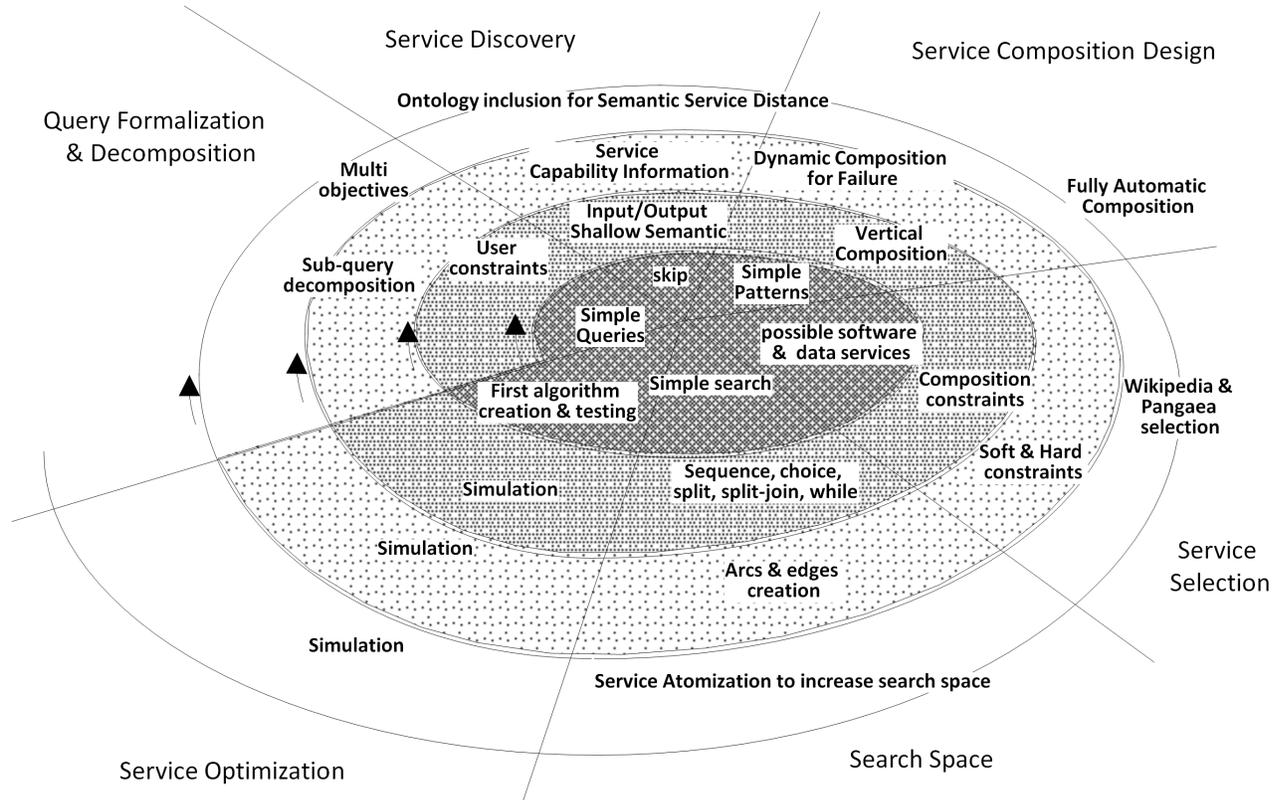


Fig. 4. Service-oriented framework for data citation system.

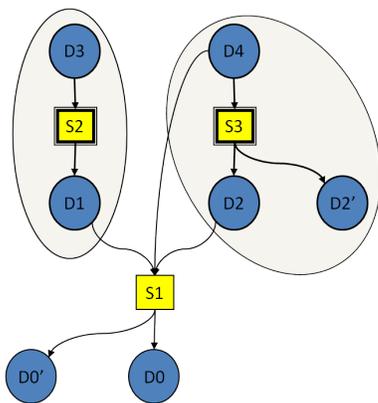


Fig. 5. Data citation service workflow 1.

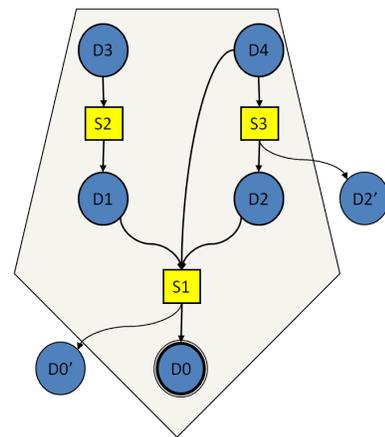


Fig. 6. Data citation workflow 2.

model inside. In data citation service architecture, both data and metadata are requested and published through various methods, such as XML, JSON, or atom publishing protocols. This architecture is proposed on top of the previous experience in combining two different architectures, such as workflow based services and pipelining based components [15].

Since this research is not only using the traditional datasets, the architecture aims at accommodating both relational data database and various data source, such as big data in Pangaea which has spatial and temporal information. Hence, the integration between the conventional database management

system and others are enabled through this architecture. The everything as a service are considered as functional datasets in order to be able to combined with relational datasets.

VI. DISCUSSION

Data citation services are openly published to the earth and environment scientists in north Africa and middle east countries. The right configuration is required to access the service. The example of the configuration of preprocessing filters and user defined index is illustrated in Fig. 8. This con-

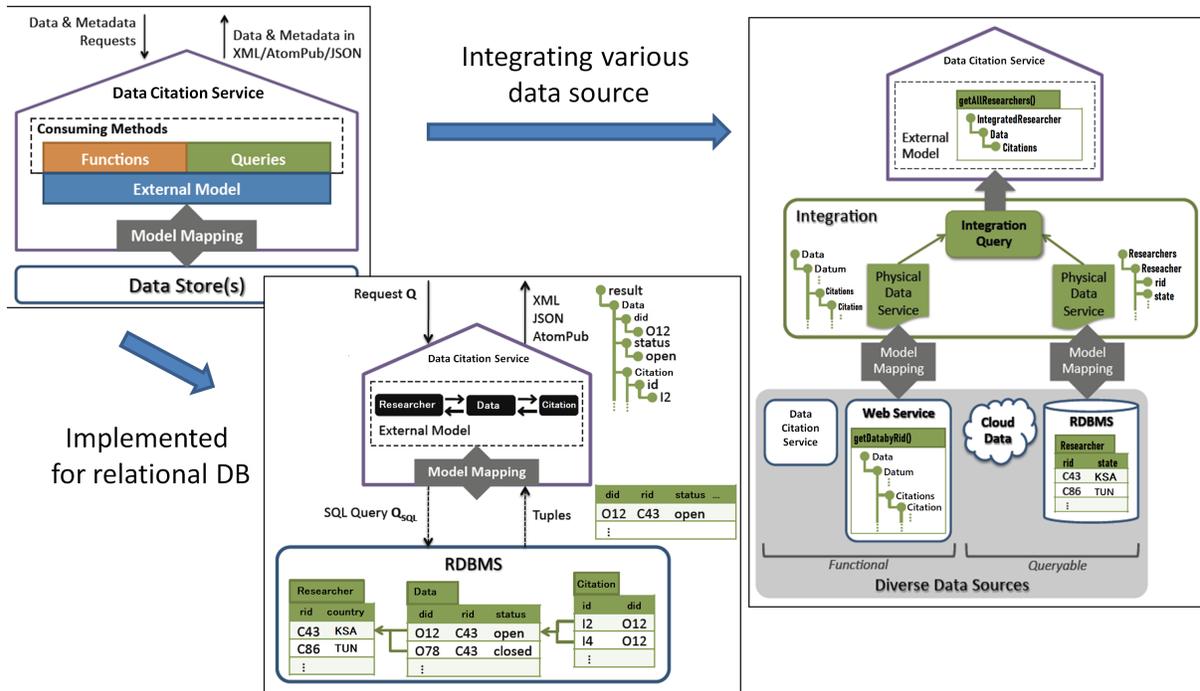


Fig. 7. Data citation service architecture.

figuration burdens the data center hardware due to the request of the dedicated resource allocation for each searching index. The structure includes HTML documents, structure analysis, parsing result, caller information, page ranking, evaluation of expression, system coordinate which includes latitude, longitude, and altitude.

percentage since there is an execution of several services in one workflow. There is a bottleneck to access the service due to the limitation of the data center hardware capability. There is also an excess of the internal university proxy to restrict the access, which is the base of future work to include a dedicated proxy to improve the access.

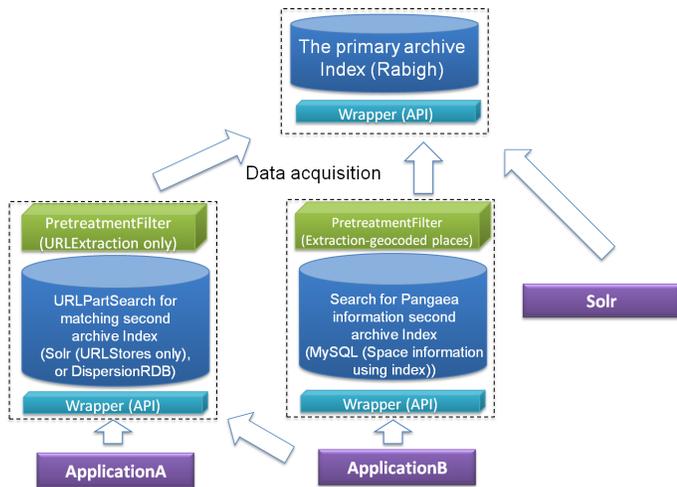


Fig. 8. Configuration example for data citation service.

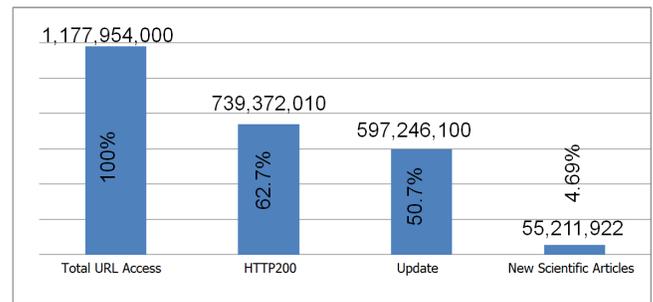


Fig. 9. Performance evaluation based on the number of access.

Fig. 9 represents the evaluation of the performance for data citation service. From total URL access, HTTP200 represents the most access by 62.7%, while updating the data citation holds 50.7%. It is interesting to note that there is a new articles related to scientific data written by the scientists through this citation enablement, although takes the portion of 4.69% only. In addition, the total access might not fully represent the total

Full bandwidth is assigned to data citation service, however not all accesses utilized fully the bandwidth. Fig. 10 represents the overall bandwidth utilization mainly on the processing time. It is interesting to note that the access of the scientific data in north Africa and middle east countries has a peak in working hours. This pattern reoccurs in any days, although there is a less amount of peak access during weekends. It infers that there is still a scientific activity beyond normal working hours, although not as much as the one during working hours.

VII. CONCLUSION

This research developed a service-oriented framework and architecture for data citation service. Through this framework

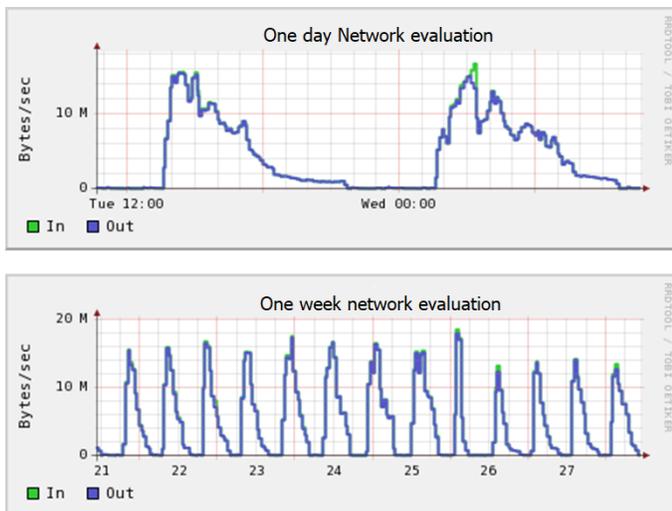


Fig. 10. Bandwidth evaluation.

and architecture, a robust mechanism is available for scientists, especially in the field of earth and environment, to relate scientific articles with scientific datasets as a big data. This framework enables the scientists; especially in Middle East and North Africa region, to cite their articles with worldwide big scientific data. In the evaluation, it is shown that the framework is robust and reliable enough to handle both big data access and its citation. It remains the future work on how to increase the relevance of the data citation, especially to attract more scientists in this region by utilizing a social media in collaborating the citation.

ACKNOWLEDGMENT

This work was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, Saudi Arabia, under grant no. 830 - 16 - D1439. The authors, therefore, gratefully acknowledge the DSR technical and financial support.

REFERENCES

[1] G. Johnson, "All science is computer science," *The New York Times*, vol. 25, pp. 1–5, 2001.

[2] M. Zhao, E. Yan, and K. Li, "Data set mentions and citations: A content analysis of full-text publications," *Journal of the Association for Information Science and Technology*, vol. 69, no. 1, pp. 32–46, 2018.

[3] S. B. Davidson, P. Buneman, D. Deutch, T. Milo, and G. Silvello, "Data citation: A computational challenge," in *Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*. ACM, 2017, pp. 1–4.

[4] S. A. El-Seoud, H. F. El-Sofany, M. A. F. Abdelfattah, and R. Mohamed, "Big data and cloud computing: Trends and challenges," *International Journal of Interactive Mobile Technologies (IJIM)*, vol. 11, no. 2, pp. 34–52, 2017.

[5] I. Foster, C. Kesselman, J. M. Nick, and S. Tuecke, "Grid services for distributed system integration," *Computer*, no. 6, pp. 37–46, 2002.

[6] J. Dean and S. Ghemawat, "Mapreduce: simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.

[7] E. Gonzales, X. Zhang, Y. Akahoshi, Y. Murayama, and K. Zettsu, "Data citation wiki for harnessing collective intelligence on document-to-data associations to interdisciplinary data access," in *23rd International CODATA Conference, Taipei, China, 2012*.

[8] Y. Murakami, M. Tanaka, A. Bramantoro, and K. Zettsu, "Data-centered service composition for information analysis," in *2012 IEEE Ninth International Conference on Services Computing*. IEEE, 2012, pp. 602–608.

[9] D. Mietchen, J. McEntyre, J. Beck, and C. Maloney, "Adapting jats to support data citation," in *Journal Article Tag Suite Conference (JATS-Con) Proceedings 2015*. Force11 Data Citation Implementation Group, National Center for Biotechnology Information (US), 2015.

[10] M. J. Carey, N. Onose, and M. Petropoulos, "Data services," *Communications of the ACM*, vol. 55, no. 6, pp. 86–97, 2012.

[11] C. Gupta and M. Govindaraju, "Framework for efficient indexing and searching of scientific metadata," in *Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*. IEEE Computer Society, 2010, pp. 553–556.

[12] H.-y. Paik, A. L. Lemos, M. C. Barukh, B. Benatallah, and A. Natarajan, "Web service composition: Data flows," in *Web Service Implementation and Composition Techniques*. Springer, 2017, pp. 193–202.

[13] A. Bramantoro, R. K. Alshammari, and A. O. Almagrabi, "Measuring business entities in multi-ontologies," *International Journal of Electronic Business*, vol. 13, no. 4, pp. 342–358, 2017.

[14] A. Bramantoro, T. Kamada, M. Tanaka, Y. Murakami, and K. Zettsu, "Towards service atomization for analyzing information," in *2012 IEEE 19th International Conference on Web Services (ICWS)*, June 2012, pp. 676–677.

[15] A. Bramantoro, U. Schäfer, and T. Ishida, "Towards an integrated architecture for composite language services and multiple linguistic processing components," in *International Conference on Language Resources and Evaluation*, 2010, pp. 3506–3511.