PROPOSING A LINKED OPEN DATA SOLUTION TO ENABLE INNOVATIVE DATA SERVICES: A CASE OF BPS-STATISTICS INDONESIA

Eko Teguh Widodo[1, 2]
[1] *Global Information and Telecommunication Technology Program, KAIST, Republic of Korea*
[2] *Information System Analyst, BPS-Statistics Agency, Indonesia*

ABSTRACT – Government data in worldwide increases significantly. To generate new values, the government publishes a web portal as Open Government Data (OGD) implementation, presenting various data fields and formats. The Government of Indonesia establishes the OGD initiative through One Data Indonesia and entrusts BPS-Statistics Indonesia as a mentor for all government agencies in Indonesia. BPS has more than 500 web portals diverged from central, provincial, and municipal representatives. In addition, data integration solutions, such as a static and dynamic table, search engine, and web API, are provided to deliver data and information for the users. However, those solutions could not respond to users' needs instead of producing various data formats (i.e., HTML, PDF, XLS, CSV, XML, and JSON), taking much time and effort to transform into information or knowledge. Linked Open Data (LOD) offers a new publishing way to reform reusability, interlinking, and sharing data on the Web. Many countries have practiced it in government sectors. Under the umbrella of the Semantic Web, those data formats and their metadata can be represented as semantic information, and LOD integrates the information, making it easily readable and consumable by humans and machines. This research aims to propose a solution using LOD for BPS to enable innovative data services. It was conducted by analyzing the as-is of data integration solutions implemented by BPS compared to government agencies in the benchmarking countries Republic of Korea and Brazil. The to-be model is also designed based on findings from the gap analysis and is expected to be executed as a pilot project in BPS. Moreover, future work should cover technology adoption in all government agencies to enable more innovative data services in Indonesia.

Keywords: BPS-Statistics Indonesia; Open Government Data (OGD); data integration; Linked Open Data (LOD); Linked Open Government Data; Semantic Web.

1. INTRODUCTION

Data integration is defined as the process of combining data from heterogeneous sources, and it can be distinguished into manual integration, common user interface, integration by applications, integration by middleware, uniform data access, and common data storage [1]. As the Non-Ministry Government Agency directly responsible to the President, BPS-Statistics Indonesia currently implements some of the approaches into concrete data integration solutions, such as a static[1] and dynamic[2] table, Allstats[3] search engine, and web API[4], to help users quickly get the desired information and knowledge. However, these solutions could not perform as users expected in immediately receiving the desired information and knowledge. It takes a lot of time and effort

[1] https://www.bps.go.id/subject/12/kependudukan.html#subjekViewTab3
[2] https://www.bps.go.id/subject/12/kependudukan.html#subjekViewTab5
[3] https://www.bps.go.id/searchengine
[4] https://webapi.bps.go.id/developer

before and after gathering data. As a result, the users will get data in various formats. Moreover, they should extract a subset of interest and manually match the data from different datasets before further processing to get insights.

Linked Open Data (LOD) is introduced as one of the solutions, described as a method to publish Open Data using Linked Data principles developed by the World Wide Web Consortium (W3C) [2]. The document-based produced from the existing solutions, including its metadata, can be represented into semantic information. Other data sources that have similar meanings; can then be interlinked and possible for reusability and sharing. Thus, new innovative data services may be generated, facilitating users to get information and knowledge quickly without spending much time and effort. Many countries have adopted LOD in the government [3-12], and BPS needs to adopt appropriate solutions.

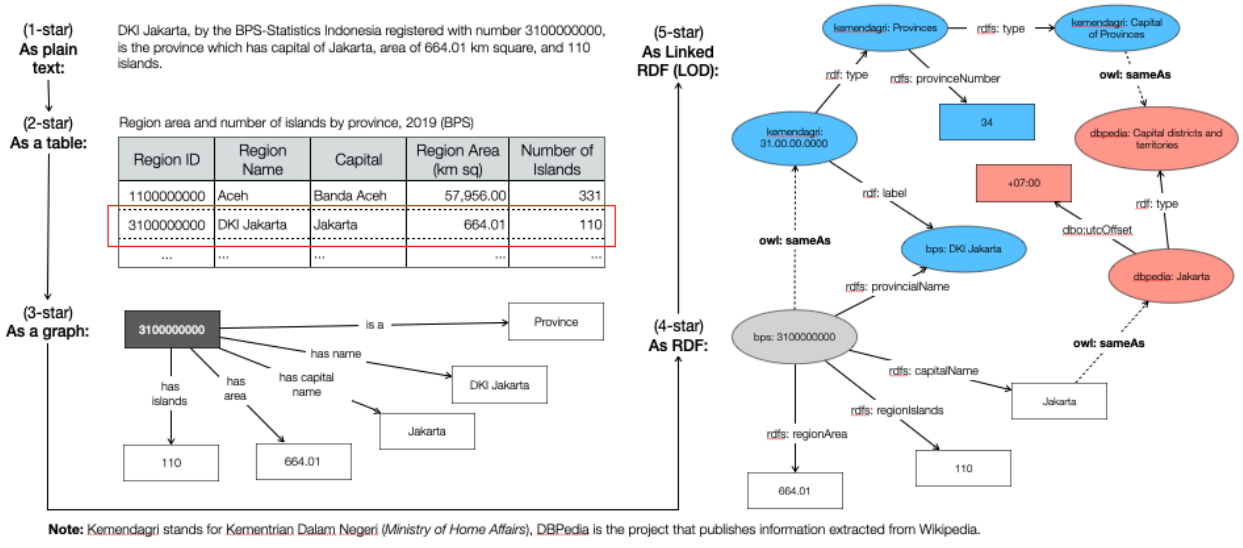
In this research, we contribute to (i) identify critical success factors (CSFs) that will be used for analyzing the situation of government agencies, (ii) analyze and compare the government agencies in Indonesia and the benchmarking countries to get findings for BPS-Statistics Indonesia, and (iii) propose a best-fit solution for BPS to adopt a LOD technology.

2. LITERATURE REVIEW

Linked Data is one of the Semantic Web technologies –*a collection of technologies and standards that allow machines to understand the meaning (semantics) of information on the Web* [13]. Publishing LOD on the Web should follow the Linked Data principles:

a. Use URIs (Unique Resource Identifiers) to name THINGS
b. Use URIs for those THINGS if possible
c. Provide helpful information at that URL
d. Include links to other THINGS

Sir Tim Berners-Lee, the inventor of the Web, suggested the 5-star level of Open Data. The higher the level of Open Data, the more data consumers and data producers benefit from it. **Figure 1** presents Open Data gathered from BPS and could be linked with other sources, such as the Ministry of Home Affairs and DBPedia.



Note: Kemendagri stands for Kementrian Dalam Negeri (*Ministry of Home Affairs*), DBPedia is the project that publishes information extracted from Wikipedia.

Figure 1. Representation of Open Data from each level

Critical Success Factors (CSFs) Identification

The concept of success in technology-driven project implementations might be diverse among the project stakeholders, where they have different expectations and evaluations [14]. Hence, the interpretations of the success will be various. We extract the CSFs from the various sources [3-12], categorized into the pioneer countries (the United States and the United Kingdom), the European, Latin American, and Asian countries. The extraction improves the data quality in technical and managerial aspects and lessons learned for government, community, and citizens/users. **Table 1** presents the CSFs identified from the findings.

Table 1. The CSFs identified from the LOD adoption in the government sectors

Government(s)	Findings	CSFs	Sources
The United States	The government/organization intention; research collaboration with the community (academia and researchers); available strategy/policy/regulation (NAP); IT infrastructure support	(1) organization intention (2) academia and research community support (3) strategy/policy, regulations, and legislation (4) IT infrastructure support (5) data integration solutions (6) prioritization (7) data and metadata (8) financial resources (9) technical framework (10) data integration solutions (11) technical framework (data, metadata, standards, URIs) (12) skills and competencies (13) level of open data (14) industrial support (15) types of interoperability (static, dynamic, syntactic, semantic)	[3], [4], [5]
The United Kingdom	The collaboration of AKT research institute (University of Southhampton) with the UK government; prioritization of data and the actor (the AKT) who run the project; some datasets (10,000) that have data and metadata; financial resources (external); technical framework; IT infrastructure; data integration solutions		[6], [7]
European countries	Collaboration, legislation, policy or strategy, and regulation; technical framework (data, metadata, standards, URIs); public body support (leadership, policy, financial, and capacity building); user participation/engagement; IT infrastructure		[8], [9], [10]
Latin American countries	Academia and research support, mainly from W3C Brazil; skills and competencies of the ICT professional officers after giving training from the W3C; improved level of open data; creating data integration solution by mashup linked-RDF		[11]
Asian countries	Top-management support; policy and regulation, academia and research support, industrial support, financial resources; organization intention; various types of interoperability (static, dynamic, syntactic, semantic)		[12], [13]

3. METHODS

After identifying the CSFs, we classify those factors using the TOE framework to explain the influence in adopting a technological innovation by an organization based on the technological, organizational, and environmental context [15]. This framework is used to analyze the as-is analysis of Indonesia by taking a case of BPS and the government agencies from the Republic of Korea and Brazil in the benchmarking analysis. The as-is analysis produces enablers (strengths) and barriers (weaknesses). The benchmarking analysis produces lessons learned. As a result, both as-is and benchmarking results are compared in the gap analysis to discover findings for BPS. In conclusion, as shown in **Figure 2**, all the results become references to construct recommendations and the to-be model.

In selecting benchmarking countries, we select a country with a top-level and another country with a middle level that Indonesia could catch up. Based on some recent vital indicators we gathered, such as ICT Development Index, E-Government Index, Open Data Barometer Index, OURdata Index, R&D Investment, and LOD adoption, we choose South Korea as the top one and Brazil as the middle one.

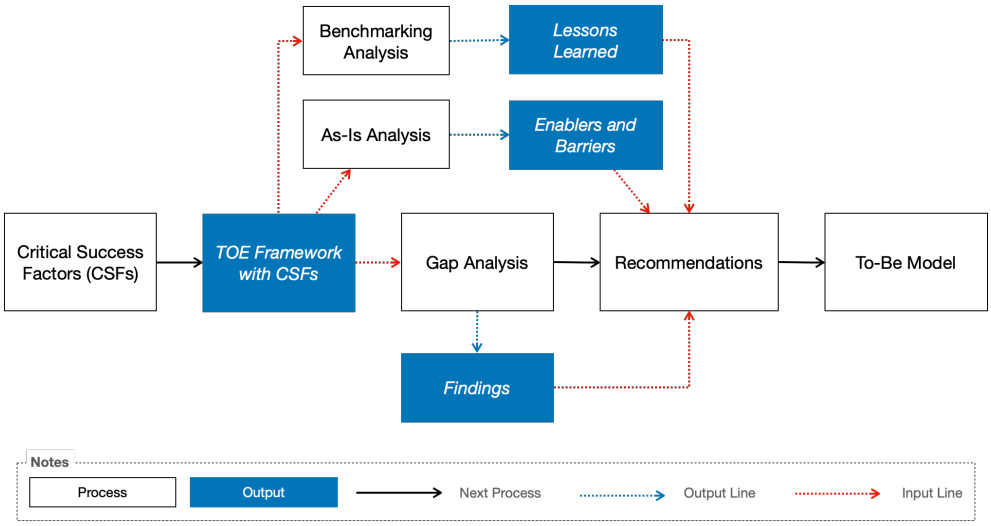


Figure 2. Research framework methodology

4. EXPECTED RESULTS

4.1. As-Is Analysis: The Case of Indonesia

As one of the founding countries that launched the Open Government Partnership (OGP) in 2011, Indonesia has committed to strengthening government data. Some strategies, policies, and regulations have been made, such as the Open Government Indonesia (OGI), One Data Indonesia (ODI), and recently Satu Data Indonesia (SDI). As one of the coaches/mentors, BPS has been selected for this research because the agency becomes a role model and data reference for all government agencies. However, although the agency has integrated data from various sources and built data services for users, there are still some lacks. The agency only provides data services that produce various data formats, which take much time and effort to obtain the information and knowledge. The data could reach 3-star of 5-star of Open Data level, where some data has been converted into CSV/JSON format, but most of them are still in the XLS format. A reported study informed that the agency has no standardized framework in improving skills and competencies of the workers and level of Open Data to use LOD technology [16]. Moreover, the agency also collaborates less with academia and the research community for research and development and promotion of data utilization. In 2020, BPS initiated the Indonesia Data Hub (INDAH) to overcome the lack and started gathering information to adopt LOD.

4.2. Benchmarking Analysis: The Case of South Korea

In South Korea, after joining the OGP in 2011, the government enacted the Public Data Act in 2013 as a foundation for managing and strengthening data quality. Moreover, other strategies and policies were established, such as creating Public Data Strategy Committee, The 1st, 2nd, and 3rd Basic Master Data Plan, Public Data Innovation Policy, and Data/AI Economic Activation Plan. Technology transfer was also made with the first movers of LOD, the United States and the United Kingdom. As a result, with the government's support, priority, and intention to quickly adopt the technology, it made some changes. Under the Ministry of Interior and Safety (MOIS) and National Information Society Agency (NIA), a collaboration with many stakeholders (i.e., research institutes, universities, private companies) was made to provide a roadmap for service platforms, data representation technology, and data provision technology. LOD is one of the roadmaps, and it has been implemented in 14 public/private organizations. At least three private local companies (e.g., LiST,

Saltlux, Frotoma) installed their innovation to those representatives to support the technological parts, like conversion data into RDF, storing to triplestores, and accessing the data using SPARQL. We study several government agencies, such as Seoul Open Data Plaza LOD using LiST technology, the Korean Intellectual Property Office (KIPO) using Saltlux technology, and the Ministry of Foreign Affairs (MOFA) using Frotoma technology. The agencies have successfully adopted and exploited it into innovative data services, like data visualization and semantic search. In addition, some events, such as the LOD conference (LODAC), are conducted to facilitate more discussion and collaboration in improving data quality and hackathon to gain data utilization and innovation.

4.3. Benchmarking Analysis: The Case of Brazil

In Brazil, we found that the intention, support, and priority from the top level of the government agency gave more impact to immediately adopt the LOD, particularly after the United States implemented it. The government enacted the regulation the Law on Access of Information (LAI) to regulate the constitutional rights of citizens to access government information and established the Information Organizing Committee of the Presidency (COI) to gather large amounts of aggregated data for digital publication. Additionally, the collaboration with the external entities, such as the W3C of Brazil and research communities, had been made, dramatically affecting the ICT professionals in the public sector to improve their skills and competencies. Furthermore, a technical framework (e.g., ePing, INE, INDA) is already provided as a guideline to enhance data quality. As a result, the Federal Budget Secretariat (SOF), as our study case, already adopted the technology regarding Open Fiscal Data and performed data visualization of budget and expenditure data and social security budget from 2000 onwards with limited IT infrastructure and developing the existing system (SIOP) into semantic information that anyone could use. To boost the utilization, the government held a contest or hackathon.

4.4. Gap Analysis

Table 2 presents the comparison of LOD implementations in the selected government agencies based on the defined CSFs. The findings of this analysis will be discussed in **Section 5**.

Table 2. Comparison by CSFs of Indonesia, South Korea, and Brazil

Context	CSFs	Indonesia	South Korea	Brazil	Findings for BPS
Technology (T)	Level of open data (maximum of 5-star)	3-star	5-star	4-star	*Low level of open data*
	Type of interoperability (static, dynamic, syntactic, semantic)	reach syntactic	reach semantic	reach semantic	*Only publishing data on the Web without adding semantic (meaning)*
	IT Infrastructure	○	✓	○	*No available IT infrastructure for LOD*
	Data integration solutions	×	✓	✓	*Existing solutions did not provide a LOD solution*
	Data and metadata	✓	✓	✓	*Well managed data and metadata using information systems*
	Technical framework (i.e., standards, license, URIs)	×	✓	✓	*No specific guideline in adopting LOD*
Organization (O)	Top-management support	✓	✓	✓	*More support, such as establish a committee or team focused on improving data quality*
	Organization's intention	✓	✓	✓	*INDAH platform as a future vision using LOD*
	Prioritization	○	✓	✓	*Less prioritization with other activities (survey, census.)*

	Skills and competencies	×	✓	✓	*Low skills and competencies in LOD*
	Financial resources	internal	internal & external	internal	*BPS only relies on internal budget*
Environment (E)	Legislation, strategy/policy, and regulations	○	✓	✓	Unclear strategy and should be synchronized
	Academia and research support	○	✓	✓	*Less collaboration with academia and research community*
	User participatory	○	✓	✓	*Less promotion on data utilization*
	Industrial support	×	✓	×	*BPS may use technology from industry or develop from scratch*

Note: ✓ (has already applied) ○ (still developing) × (has not applied yet)

4.5. Recommendation and To-Be Model

Based on the enablers and barriers from the as-is analysis, lessons learned from benchmarking analysis, and findings in the gap analysis, we construct the recommendations for BPS. There are five significant recommendations: (i) create a LOD strategy, (ii) collaborate with stakeholders, (iii) provide capacity building/training, (iv) transfer the technology, and (v) build a LOD system. In this research, we focus on proposing a technical solution regarding adopting the LOD in BPS. Therefore, as shown in **Figure 3**, we build a LOD system.

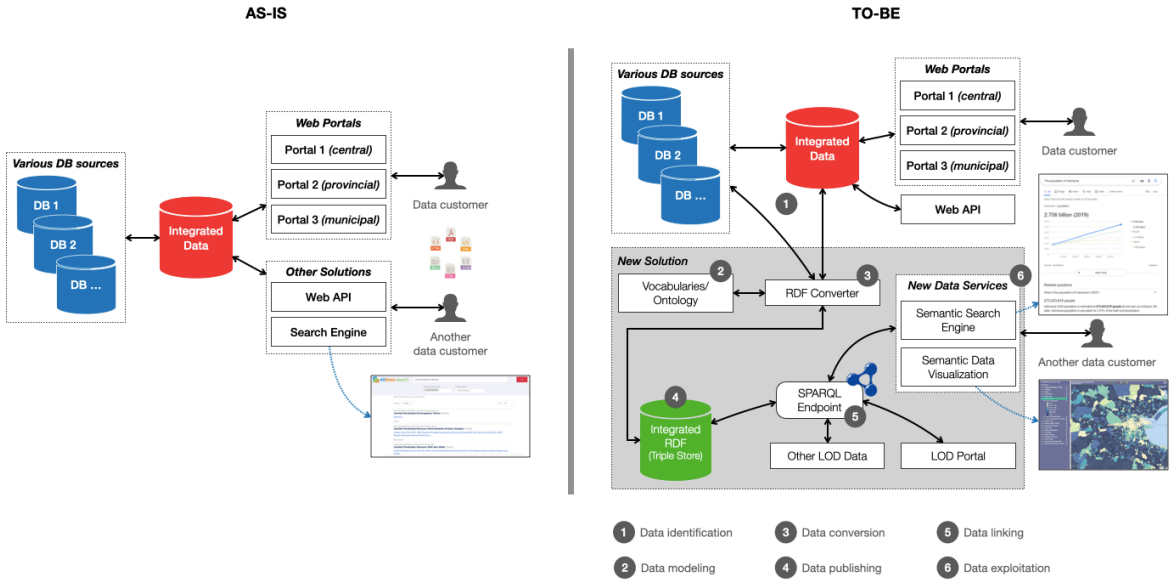


Figure 3. Proposed a LOD solution compared between the as-is and to-be

As shown in **Figure 3**, we categorized the solution into six stages. The data identification stage selects and analyzes the existing BPS data sources. We can prioritize the datasets that have reached 3-star, essential domain knowledge (i.e., population, poverty, SDGs), and completed metadata. In addition, we could also define the URIs and set the license. In the data modeling stage, we use some tools to design data models using ontology engineering. The data conversion stage will convert the results from the first two stages into an RDF data format. In the data publishing stage, those RDF will be stored in a data storage system called triplestores. Then, we also provide a system to link with other RDF sources in the data linking stage. Finally, in data exploitation, we could perform data utilization from Linked-RDF (LOD) to build an innovative data service that helps the users get the information or knowledge fast, such as Semantic Search Engine and Semantic Data Visualization.

5. DISCUSSION

This research aims to identify the CSFs to adopt the LOD in the government agency. We have already found the factors. However, some drawbacks need to improve. The first is to determine which factors influence the adoption by conducting quantitative research through a survey of stakeholders. Additionally, some sources seem not quite up-to-date. It is around 2011-2014, after the OGP was launched. Thus, strengthen the influence factors and collect more recent data would be our future work to provide a best-fit solution for BPS in adopting the technology.

Another goal of this research is to get findings for BPS after analyzing and comparing the government agencies in Indonesia, the Republic of Korea, and Brazil according to the defined CSFs. We have found that BPS has some lacks comparing to benchmarking countries. The agency mainly lacks establishing an IT infrastructure and technical framework to adopt LOD in a technological context. In the organizational context, BPS has less support from top-level management, intention, and priority to seriously adopt this technology through actions, such as establishing organizational structure (team or committee), improving skills and competencies. BPS also lacks collaboration or partnership with academia, the research community, and industry to enhance innovations and data utilization for environmental context.

Lastly, we also construct the recommendations and propose a new solution for data integration in BPS. We have defined the recommendations in **Sub Section 4.5** and focused on the solution by building a LOD system. Thus, there are six stages to develop. In the data identification stage, we identify and analyze existing data and metadata. We strongly recommend using 3-star datasets to be processed. For the data modeling stage, we can model the schema from existing metadata. Then, the following stage is data conversion, where data and its schema are converted into RDF data. In the data publishing, those RDF data is stored in the data storage called "triplestores." The data linking stage will link the RDF within the storage or external storage to become Linked-RDF or LOD data. Finally, we can exploit those LOD to create innovative data services, such as semantic search engines or data visualization.

To sum up, we have discussed that this research has reached its goals. However, there are still some limitations that need to be improved in enhancing research quality. Even though the research concentrates more on building a LOD system that describes a technical thing, it should discuss other recommendations to get comprehensive information to build the system. Furthermore, it can be implemented as a project for BPS and adopted by other government agencies.

6. CONCLUSION

This paper successfully presents the CSFs identified from the LOD projects of government agencies worldwide. In addition, it also found the findings for BPS after comparing those factors to the Republic of Korea and Brazil. Besides, as part of recommendations, we focus on the technical solution by establishing a LOD system consisting of six stages, starting from data identification, data modeling, conversion, publishing, and exploitation. Each stage may use technologies that best-fit for BPS. Despite those, continuous research is required to overcome the pitfalls, such as quantitative research to evaluate the CSFs and non-technical research for other defined recommendations.

AUTHOR CONTRIBUTIONS

The author conceived and designed the analysis, collected the data, contributed data or analysis tools, performed the analysis, and wrote the paper.

REFERENCES

1. Ziegler, P., and Dittrich, K.R. (2007). Data Integration - Problems, Approaches, and Perspectives. In: Krogstie J., Opdahl A.L., Brinkkemper S. (eds) Conceptual Modelling in Information Systems Engineering. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-72677-7_3
2. Sakr, S., Wylot, M., Mutharaju, R., Phuoc, D.L., and Fundulaki, I. (2018). Linked Data: Storing, Querying, and Reasoning. Springer. ISBN: 978-3-319-73514-6. https://doi.org/10.1007/978-3-319-73515-3
3. Hendler, J., Holm, J., Musialek, C., and Thomas, G. (2012). US Government Linked Open Data: Semantic.data.gov. IEEE Intelligent Systems, 27(3), pp. 25-31. https://doi.org/10.1109/MIS.2012.27
4. Ding, L. et al. (2011). TWC LOGD: A portal for linked open government data ecosystems. Journal of Web Semantics, 9(3), pp.325-333. https://doi.org/10.1016/j.websem.2011.06.002
5. Gramaglia, L., Kormann-Fromageau, C., Delcambre, D., Museux, J. M., and Nagy-Rothengass, M. (2018). A European strategy for linked open statistics. CEUR Workshop Proceedings (Vol. 2317). CEUR-WS. http://ceur-ws.org/Vol-2317/article-11.pdf
6. Lebo T. et al. (2011). Producing and Using Linked Open Government Data in the TWC LOGD Portal. In: Wood D. (Eds) Linking Government Data. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-1767-5_3
7. Shadbolt, N. et al. (2012). Linked Open Government Data: Lessons from Data.gov.uk. IEEE Intelligent Systems, 27(3), pp. 16-24. https://doi.org/10.1109/MIS.2012.23
8. Shadbolt, N., and O'Hara, K. (2013). Linked Data in Government. IEEE Internet Computing, 17(4), pp.72-77. https://doi.org/10.1109/MIC.2013.72
9. Deirdre, L. (2014). Building an open data ecosystem: an Irish experience. Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance (ICEGOV '14). Association for Computing Machinery, New York, NY, USA, pp.351-360. https://doi.org/10.1145/2691195.2691258
10. Breitman, K. et al. (2012). Open government data in Brazil. IEEE Intelligent Systems, 27(3), pp.45-49. https://doi.org/10.1109/MIS.2012.25.
11. Kim, H. (2018). Interlinking Open Government Data in Korea using Administrative District Knowledge Graph. Journal of Information Science Theory and Practice, 6(1), 18–30. https://doi.org/10.1633/JISTAP.2018.6.1.2
12. National Information Society Agency (NIA). (2014). 링크드 오픈 데이터 국내 구축 사례집. Retrieved from http://lod.seoul.go.kr/home/guide/2014_Linked_Open_Data_internal_build_case.pdf [Online; accessed: 21-April-2021].
13. Allemang, D. and Hendler, J. (2011). Semantic Web for the Working Ontologist 2nd Edition: Effective Modeling in RDFS and OWL. Morgan Kaufmann. ISBN: 978-0-123-85966-2
14. Zuiderwijk, A., Susha, I., Charalabidis, Y., Parycek, P.,a nd Janssen, M. (2015). Open data disclosure and use : critical factors from a case study. CeDEM 2015 : Proceedings of the International Conference for E-Democracy and Open Government, p197-208. http://urn.kb.se/resolve?urn=urn:nbn:se:oru:diva-50203
15. Baker J. (2012). The Technology–Organization–Environment Framework. In: Dwivedi Y., Wade M., Schneberger S. (Eds.), Information Systems Theory. Integrated Series in Information Systems, vol 28. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-6108-2_12
16. Rahmatika, M., Krismawati, D., Rahmawati, S.D.,, Arief, A., Sensuse, D.I., and Fadhil Dzulfikar, M. (2019). An Open Government Data Maturity Model: A Case Study in BPS-Statistics Indonesia. The 7th International Conference on Information and Communication Technology (ICoICT), 2019, pp.1-7. https://doi.org/10.1109/ICoICT.2019.8835352
17. Romzi, M. (2020). Orkrestasi Satu Data Indonesia dengan INDAH (Indonesia Data Hub). Presentation Material, Jakarta, June 18th, 2020.
18. BPS Ontology Team. (2020). Dukungan Ontologi terhadap Indonesia Data Hub versi 1.2. Presentation material. BPS-Statistics Indonesia
19. National Information Society Agency (NIA). (2014). Public Database Service Technology Trend, 데이터베이스 활용기술 전망. Retrieved from https://www.slideshare.net/jino/ss-63404561 [Online; access: 21-April-2021].
20. Federal Budget Secretariat (SOF). SIOP - Dados Abertos. Retrieved from https://www1.siop.planejamento.gov.br/siopdoc/doku.php/acesso_publico:dados_abertos [Online; accessed: 10-April-2021]