UNDERSTANDING BARRIERS IN THE IMPLEMENTATION OF THE ONE DATA POLICY IN INDONESIA: INSIGHTS FROM HEALTH DATA JOURNEY MODELLING

Arthur Glenn Maail
Open Data Lab Jakarta

Abstract *– The Indonesian One Data policy is designed to improve internal government data governance practices by providing a regulatory framework concerning organizational structure, including the roles, tasks, and responsibilities of each key stakeholder. It also specifies mechanisms to ensure the preparation, collection, and/or processing of data that meets data standards, the application of metadata according to the standard format, and dissemination of data according to the principles of data interoperability. We conducted a data journey modelling for three key health datasets to identify challenges and barriers in data flow across local and national government agencies. The findings highlight the critical role of the local government leaders and data custodians, enforcement of data standards and policies, and compliance (including a mechanism to enforce penalties for non-compliance) to the successful implementation of the One Data policy in Indonesia.*

Keywords – Data governance, data journey modelling, data standard, Indonesia, local government

1. INTRODUCTION

One Data Indonesia (One Data) is an initiative of the Government of Indonesia (GoI) to improve interoperability and the use of government data. It is envisioned that the accuracy, openness, and interoperability of government data will be assured by improving data governance practices through structuring regulatory and institutional frameworks, standardizing and synchronizing data assets, building the capacity of agencies and government instrumentalities, and facilitating data interoperability across government agencies. At the same time, the One Data policy seeks to improve data disclosure practices by ensuring the availability of high-quality data for the public through open data portals at both national and subnational levels.

2. LITERATURE REVIEW

The recent advancement in big data technologies and applications including machine learning and artificial intelligence has spurred optimism within the government that these technologies can enhance their ability to serve the citizens and address major national challenges involving the economy, healthcare, job creation, natural disasters, and terrorism [1, 2]. From "smart" government to transformational government, Big and Open Government Data can foster collaboration, promote greater openness and usher in a new era of policy and decision making [3].

However, before any big data program can produce a meaningful impact, prior studies have highlighted an importance of addressing a range of policy challenges concerning data governance including a robust technology infrastructure for organizing, curating, storing, and making datasets accessible among agencies and to the public as the main prerequisite [1, 3, 4]

According to DAMA Data Management Body of Knowledge (DAMA-DMBOOK), the function of data governance covers the exercise of authority, control and shared decision making (planning, monitoring and enforcement) over the management of data assets. Data governance plays a very important role in achieving high data quality. Data governance creates an organizational structure that develops and enforces policies, rules, processes, and procedures to ensure and improve data quality within an organisation.

Thompson, Ravindran [5] suggest three key aspects for effective public-sector data governance: people (leadership and data steward), standards (policies and requirements), and compliance (including a mechanism to enforce penalties for non-compliance). This study aims to contribute to the current understanding of these key aspects by investigating barriers in the implementation of data governance policy in Indonesia.

3. THE ONE DATA POLICY

The Indonesian Government has recognized the potential significance of improving internal administrative operations, including data governance. While the Indonesian Government has made efforts to advance its E-Government system, the progress is still far from satisfactory. In the 2016 United Nation's E-Government Development Index shows Indonesia is ranked at 116 out of 193 countries, behind several other south-east Asian countries and also the regional average. Much of local government is still at the initial developmental stage, providing only government information such as organization structure, contact details, news, and events via their official websites. The annual E-Government Indonesia ranking by the Ministry of Communication and Informatics shows that the top three provinces are all located on the well-developed Java island, indicating a strong dependency between availability of technology infrastructure with public sector innovations.

With regard to data governance, the report issued by the Executive Office of the President shows that at present, there are no standard data management practices applied across government agencies [7]. Each agency is working in silo, developing its own data management practice, with a lack of a clear strategy for data sharing and collaboration. The One Data policy is expected to provide a common framework for data management along with guidelines for public institutions to limit redundant efforts, improve data quality, interoperability and integration, including data licensing and formats. The roadmap for the implementation of the One Data began with the pilot implementation at nine national ministries in 2016 and 2017 [7]. In the next phase, the implementation will continue to cover all national and subnational agencies.

To support preparation for implementation at the subnational government level in 2019, the One Data Team at the Executive Office of the President and the Open Data Lab Jakarta commissioned a study to assess readiness to implement One Data provisions. The central question addressed by this research is what do local governments need in order to implement or align themselves with the anticipated One Data Policy? To address this question, we conducted data journey modelling in four municipalities in Indonesia: Pontianak , Bojonegoro, Makassar, and Kulon Progo. The following section details the methodology of the study.

4. PROPOSED METHODS

Data journey is a tool aiding the identification of social and technical barriers to data movement in large, complex organizations [8]. The tool offers an ability to model the necessary parts of the existing information infrastructure, including the places where data is stored, and the links between these places that enable the sharing of data. It can also be used to describe the movement of data from a point of entry in the infrastructure to the point of use by a new consumer (human user or a new IT development). Once a diagram of the movement has been created, it is possible to identify parts of the movement that, because of some type of barrier, can introduce some costs to the movement or new development. Hence, data journey modelling can assist the decision-making process of whether it is worth implementing a new functionality or policies on an existing network of data. The diagram contains several elements as shown in Figure 1. These elements include:

- Container or place where data is stored. A container can be in electronic form (e.g. a database, an Excel spreadsheet, a word document) or in physical form (e.g. cabinets, desks, pigeon holes). Electronic data containers are denoted using the database icon and physical ones with a rectangular box.
- Journey leg is an established route where data stored in a container can travel to another container. The direction of the link shows the movement of data from the source to the target container. The link is denoted with a straight line connecting the two containers.
- A medium is used to move data from a source container to a target container. Media can be of physical (paper) or electronic form (email, file transfer).

- An actor is a person or IT system that interacts with the container. The actor may create, consume or transform the data resting in the container. The actor is denoted using the actor symbol of the UML notation and the interaction with the containers is shown with a dotted arrow beginning from the actor and ending with the container with which he or she interacts. Several actors can interact with one container. One actor can also have several interactions with different actions with the same container.

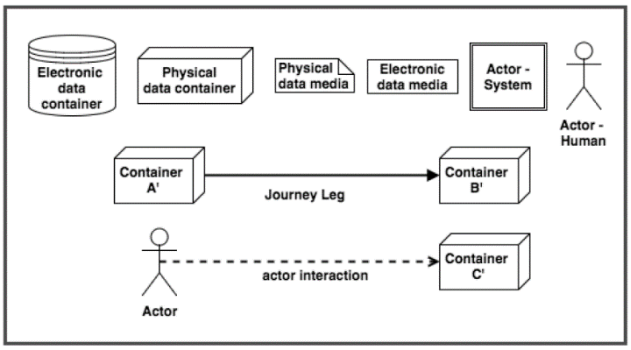

Fig. 1 – The notation used in modelling of a data journey

5. DATA COLLECTION

After consultation with the One Data team during the preparation stage, three key health datasets in the Sustainable Development Goal (SDG) indicators were selected for the assessment. They include data on (1) the number of deliveries in healthcare facilities, (2) community-led total sanitation, (3) prevalence of stunting, height for age. In this study, we modelled the flow of these datasets from the point of entry at the Community Health Centre until their end journey at the Health Ministry.

Data journey modelling began with a process of identifying organisations and actors associated with the journey for each dataset. Preliminary interviews were conducted with officials from the Ministry of Health to understand the flow of each dataset, and how it interacts with organisations, and actors in the data value chain. This was later validated during the field visit to the health department at the provincial and municipal level as well as at community health centers as shown in the organisation structure in Fig. 2.

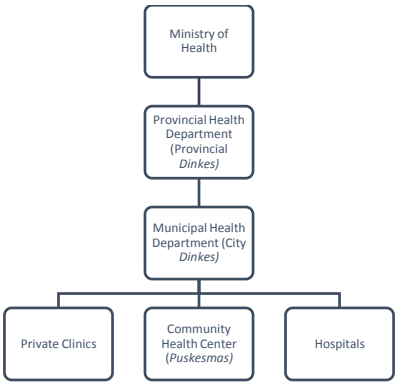

Fig. 2– Organisation structure

The study utilized a qualitative approach, in which in-depth interviews and observation of key informants responsible for data management at each level of the Health Department organisation structure were the primary methods for data collection. These include meetings with the head of department, data producers, data custodians, and IT staff. In total between 25-30 informants with the knowledge of the data journey for the three key datasets were interviewed in each city. To corroborate the insights provided by the interviews, a comprehensive desk review of city government policies, commitments, and projects related to aspects of the One Data policy was also used as a secondary source of data.

6. FINDINGS

Data on the number of deliveries in healthcare facilities are sent quarterly to the Health Ministry by the provincial health department. However, the data at the provincial level is being updated on a monthly basis by the community health center/private clinics/hospitals and subsequently by the municipal health department.

Data collection in the community health center (*Puskesmas)* is performed in two ways. Childbirth records kept by independent midwives are sent to *Puskesmas*, while registers in hospitals and private clinics are fetched manually by the *Puskesmas* midwives (leg 1). Data from both sources are combined in the District Monitoring Report book, which tis hen converted to both electronic and printed format; the latter is kept in the data folder. The printed data is delivered physically to the Maternal & Child Program office at the municipal health department (leg 2) using the format provided by the municipal health department. The data is then stored in two places; in the officer's computer and in the data folder. Citywide data is

then combined and sent to the provincial health department via e-mail (leg 3). Here, aside from being kept for the tri-monthly report to the Ministry(leg 4), data is also transferred to the Information & Data Center (*Pusdatin*), which will process the data for the province's annual health profile and for the Health Ministry data repository KOMDAT.

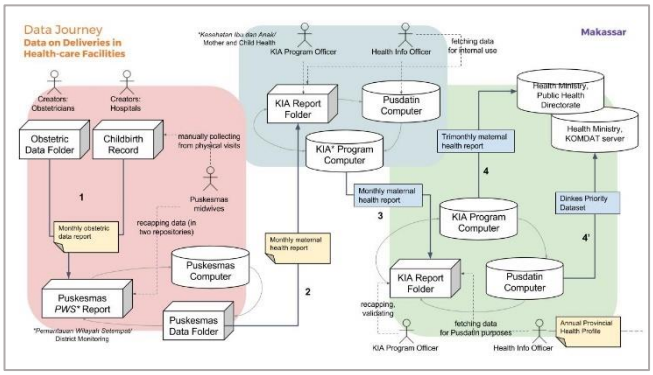

Fig. 3 – The journey of the dataset on deliveries in healthcare facilities

Table 1 shows the list of barriers identified along the legs of the journey. *Puskesmas*, as the most immediate data producers, often find the activity of collecting data from other sources most challenging. This is due to the absence of obligation for private practices and hospitals to cooperate with *Puskesmas* nearest to their area in providing data on deliveries. As a result, each month *Puskesmas* midwives must go around these locations to fetch and sort out information manually (barriers on Leg 1). Moreover, as the popularity of each *Puskesmas* in the city varies, midwives working in busier *Puskesmas* often find it harder to juggle between their main job of serving patients and processing data.

The dispersed sources of data also affect the control of information in the municipal health department. As they too receive submissions from public hospitals, the Maternal & Child Officer must host a monthly gathering of Puskesmas' officers to ensure the distribution of this data. Meanwhile, the need to convert data formats (i.e. from physical to electronic, and from one table to another) in both the city and provincial health departments also results in inefficiency in the monthly report preparation.

Table 1 – Barriers in the journey of the dataset on the number of deliveries in healthcare facilities

Journey leg number	Barrier	Cause of barrier
1	The repeated loss of data of deliveries outside *Puskesmas*	The lack of a systemized data collection procedure for hospitals and obstetric gynaecologists (ob-gyns), especially the private ones. At the moment they are not strictly required to report delivery records to either the municipal health department or *Puskesmas* and are not aware of the importance of data integration.
1	Time inefficiency in data collection	Competing tasks by the midwives. The need for *Puskesmas* midwives to physically visit hospitals and ob-gyns, sort each delivery data by address, and group each birth according to the *Puskesmas*' working boundary area.
1	Extra burden of carrying out data entry for different applications	The lack of incentive to reward officers carrying out the 'additional job' of data entry.
2	Extra task for municipal health department officer to sort out data from all community health centers, private clinics, and hospitals	The flow of data collection from hospitals, ob-gyns, and practices outside *Puskesmas* is not clear and binding.
3	The inefficiency in preparing and delivering data in both physical and electronic forms	- The Ministry and other data users require data to be processed into different tables and documents - Trust for digital registry formats among Health department staff is low, so diminishing physical formats altogether is not feasible

Data on community-led total sanitation (CLTS) are delivered vertically by two means: directly to a server at the Ministry via the Smart-STBM (*Sanitasi Total Berbasis Masyarakat*) application and manually using monthly and tri-monthly report documents.

The data input in the Smart-STBM application is done by sanitarians at the village level. Data is sent through either via the app or SMS to the STBM server (leg 1). The data is updated in real time, and accessible by both the Environmental Health Directorate in the ministry and the Provincial Coordinator in the provincial health department. After obtaining feedback from the ministry (leg 2) and aggregating the data themselves, the Provincial Coordinator relays it to municipal health department officers as recommendations (leg 3), who then do the same to *Puskesmas* officers (leg 4). For instance, feedback from the top may include an instruction to focus on certain areas or aspects of sanitation. This way, sanitation data on the STBM application is always kept updated.

However, interviews done with *Puskesmas* officers revealed that data transacted through the STBM application is only available for the first of the five CLTS pillars (i.e. Stop Open Defecation), thus requiring another way of reporting for the remaining pillars. The manual delivery consists of sanitarians sending printed reports to the sanitation officer in *Puskesmas* (leg 1') which then summarize the data before delivering it to the municipal health department (leg 2'). The data from all *Puskesmas* is then processed in a fairly similar way as data on Number of Deliveries and Stunting, which is bottom-up to the provincial health department (3'). At this point, data will be then converted into either a tri-monthly report for the Environmental Health Directorate (4') or the province's annual health profile.

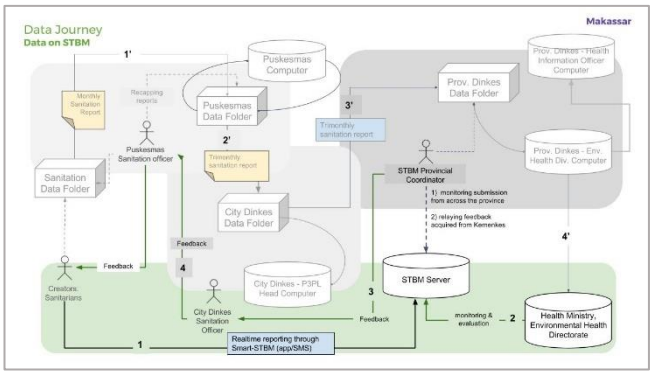

Fig. 4 – The journey of the dataset on community-led total sanitation

Table 2 shows the barriers encountered by the producers and users of STBM data. The introduction of the application-based data entry to help management and monitoring proves to be beneficial in keeping the timeliness of the data. However, barriers related to technology still causes delays at the lower level, which requires a provincial CLTS coordinator to directly resolve the issue.

The application also does not cover data input for other CLTS pillars outside Open Defecation. Therefore, the provision of data on the other four pillars must be done physically.

Table 2 – Barriers in the journey of the dataset on community-led total sanitation

Journey leg number	Barrier	Cause of barrier
1	Inefficiency in preparing sanitation data on different formats.	The Smart-STBM app only covers data log from one out of the five pillars included in the CLTS program blueprint, which is Open Defecation Free (ODF).
1	The delay in sanitarians to send off real-time information through the Smart-STBM app.	Different understanding of STBM app by sanitarians and/or technical issues (e.g. lack of signal or smartphone not working)
2'	The need to transact sanitation reports, either monthly or tri-monthly, manually outside the Smart-STBM app.	The Smart-STBM app only covers data log from one out of the five pillars included in the CLTS program blueprint, which is Open Defecation Free (ODF). The other four pillars, therefore, are sent from *Puskesmas* to the municipal health department through manual forms.
3	The need for provincial CLTS coordinator to regularly check each sanitation agent who has low-performance level according to the internal application system.	Different level of proficiency in using STBM app in the lower hierarchy and/or technical issues (e.g. lack of signal or smartphone not working).

Journey leg number	Barrier	Cause of barrier
4	The need for sanitation officers to constantly check Smart-STBM app for change in data.	No notification system and updates when there is a change in data.

Meanwhile, the production of stunting data employs two different methods: sampling and direct measurement. The first is performed annually by Nutritional Status Monitoring (*PSG*) enumerators and the provincial health department, while the direct measurement is done during monthly patient visits at Posyandus.

Data held by *PSG* enumerators, including stunting data, is first distributed to both the provincial and municipal health departments (leg 1). After the verification/'data contrasting' process is done on their level, the data is sent back to the province (leg 3) before being compiled into a publication and sent to the Health Ministry (leg 4).

Meanwhile, data produced from direct measuring is collected monthly from Posyandus (a monthly clinic for children and pregnant women) by *Puskesmas* nutrition officers (leg 1'). Here, the data gathered is converted several times across both physical and electronic formats. The *Puskesmas* officer first logs Posyandu reports to their *cohort* registry and transfers it to *PosyanduQu*, a spreadsheet application introduced by the Health Ministry. To prepare the monthly report, the numbers aggregated from the program are then moved to a nutritional book and again translated into a spreadsheet form prepared by the municipal health department. This information is then sent physically to the municipal health department (leg 2') and processed to be used for citywide health policymaking.

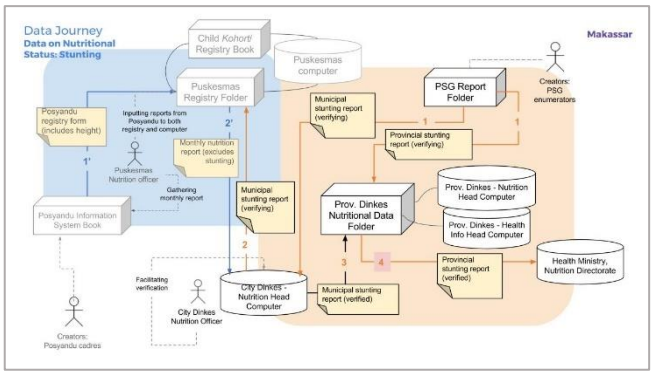

Fig. 5 – The journey of the dataset on stunting

The existence of two different methods for collecting stunting data causes several recurring issues (see Table 3). As *PSG* annual enumeration is conducted by the provincial health department under the direct supervision of the Ministry of Health, its result is what is normally used to measure a certain area's nutritional condition. Based on the PSG data, the provincial health department will then provide direct recommendations to the municipal health department and *Puskesmas*. However, this has sometimes proven to be problematic as the sampling result many times contradicts the data held by all *Puskesmas* (see barriers on journey leg 2).

Barriers on legs 2' and 2 also show that confusion happens in the municipal health department, which is tasked to deal with the processing and managing of data from the two sources. That is, the PSG recommendation from the province's health department and the direct measurement report from *Puskesmas*. Moreover, the importance of nutritional data for national, provincial, and city level policies imposes a burden on *Puskesmas* officers, who are required to convert the data into three separate formats (see barrier in journey leg 1').

Table 3 – Barriers in the journey of the dataset on stunting

Journey leg number	Barrier	Cause of barrier
1'	Issues in monthly data collection among Posyandu cadres	- Limited capacity among Posyandu cadres in producing correct measurements according to health standards. - The loss of data from toddlers who are not consistently coming to Posyandu
1'	Time inefficiency in data entry	One officer at *Puskesmas* must do data entry in several formats; including a physical register for *Puskesmas* use, *PosyanduQu* electronic spreadsheet, nutritional book, and electronic form for the municipal health department
1'	Officers juggling between main service tasks and data management	The absence of incentive to reward officers carrying out the 'additional task' of data management
2'	Difficulty in reconciling findings from two different data sources	Two different data methods, sampling and direct measurement, produce different understandings of an area's condition
2	Confusion about data flow	The current data flows require the municipal health department to process both data from the lower level (*Puskesmas*) and upper level (provincial health department)
4	The confusion created by the changing of report formats	The presence of different formats in filling in the nutritional report, including the newly introduced e-PPGBM (Community-based Registration of Nutritional Data), which requires in-depth information of each person recorded.

7. DISCUSSION

Good data governance policies and procedures are invaluable in the management of data and information in organizations. They are necessary to encourage the efficient use of human, financial and technical resources and accountability for the use of those resources. The following subsections discuss the additional insights added to the existing understanding of the three key aspects for an effective public-sector data governance in light of this study.

7.1 People

For the success of a data governance program and to provide for strategic deployment of resources, it is essential for the organization's leadership to be inspired, committed, and visionary. These leaders must not only understand the vision but be able to communicate the vision throughout the organization and motivate the stakeholders and data stewards for effective data governance. Hence, there is the need to build data governance capacities in local government (both elected and career civil servants) so that they do not only own the process of data management but will also act as champions in the implementation of the policy.

This research shows that while a strong commitment from local leadership is key, proactive engagement from national government agencies also reinforces local commitment. The commitment from the city mayor and relevant departments in implementing various data management initiatives over the last two years was made possible through constant dialogue between city government officials and central government representatives.

7.2 Standards

Similarly, the existence of a government-wide data standard is needed to create a uniform transactional and analytical environment for compliance monitoring, preventing each government entity from working in silo, developing different data standards. Data standards enable automation of data quality control processes with metadata-driven enhancement and best practices for improving data quality.

This study shows that the new policy needs to establish government-wide policies for data gathering, cleaning, and storing including by the non-government data producers such as private clinics and hospitals. These policies also need to address the mechanism of data sharing including data interoperability among different levels or units of a single governmental agency or inter-government. Also, the One Data policy needs to address the mechanism to avoid data duplication within a single governmental agency. With the

absence of a data custodian at the provincial, and city/district government levels, it is not clear how this oversight function may be exercised.

The findings also call for improving mechanisms to ensure data accuracy and completeness by data producers. The current practice shows that this task is being routinely done by data producers. However, the existence of non-machine-readable data formats creates inefficiency in data entry time and a lack of data accuracy. A monthly data updating process generally took approximately three weeks to complete. Therefore, utilisation of ICT applications that can shorten data entry time and minimise errors for data transfer to electronic formats needs to be encouraged. In the same way, the transition from paper-based reporting to digital formats needs to be initiated and carefully managed. Local government also needs to review the different information systems used in the different hierarchical levels to ensure uniformity in standards and related implementation.

7.3 Compliance

Finally, organizations may have a data governance policy and structure, but may not focus on the execution of these data policies. The proposed data standard needs to be strictly enforced with possible penalties for non-compliance.

Hence, it is critical to develop a strategy for regulatory compliance and adherence to determined policies, including incentives for a data custodian and data producers. Incentives for a data custodian and data producers should be integrated as part of the employee's key performance indicators and they also need to be taken into consideration by the data governance steering committee and regulated by laws/regulations in each agency.

The Government of Indonesia needs to review existing laws/regulations concerning tasks, roles, and authorities of data custodians and data producers in each agency. including their incentives to ensure that data management tasks do not pose an undue burden to the health service delivery activities, compromising quality in both areas.

8. CONCLUDING REMARKS

The advancement of digital technology makes it easy to create, transmit, store, access, and use information that is critical to supporting a data governance strategy, in government organisations. Develop organisation-wide data architecture standards including ensuring that metadata is available for each type of dataset, as well as procedures to maintain data accuracy and the responsibilities matrix associated with these tasks with possible penalties for non-compliance.

However, current technologies and standards have become so ubiquitous that they alone no longer provide a distinguishing competitive advantage for an organisation. It is rather people's creative use of information that counts rather than the technology alone. The leadership and vision of city leaders are crucial for ensuring a successful implementation of reform. Review of existing laws/regulations concerning tasks, roles, and authorities of data producers and data custodians, including their incentives to ensure that data management tasks do not pose an undue burden on public service delivery activities, compromising quality in both areas.

REFERENCES

[1] Kim, G.-H., S. Trimi, and J.-H. Chung, *Big-data applications in the government sector.* Communications of the ACM, 2014. **57**(3): p. 78-85.

[2] George, G., M.R. Haas, and A. Pentland, *Big data and management.* Academy of management Journal, 2014. **57**(2): p. 321-326.

[3] Bertot, J.C., et al., *Big data, open government and e-government: Issues, policies and recommendations.* Information Polity, 2014. **19**(1, 2): p. 5-16.

[4] Chen, Y.-C. and T.-C. Hsieh, *Big data for digital government: opportunities, challenges, and strategies.* International journal of public administration in the digital age (IJPADA), 2014. **1**(1): p. 1-14.

[5] Thompson, N., R. Ravindran, and S. Nicosia, *Government data does not mean data governance: Lessons learned from a public sector application audit.* Government information quarterly, 2015. **32**(3): p. 316-322.

[6] Zhang, J., *Operationalizing Data Quality through Data Governance*, in *Data Governance: Creating Value from Information Assets*, N. Bhansali, Editor. 2014, CRC Press, Taylor & Francis Group: Boca Raton. p. 65-92.

[7] UKP4, *Cetak Biru Satu Data Untuk Pembangunan Berkelanjutan*. 2014, Unit Kerja Presiden Bidang Pengawasan dan Pengendalian Pembangunan (UKP4): Jakarta.

[8] Eleftheriou, I., S.M. Embury, and A. Brass. *Data journey modelling: predicting risk for IT developments*. in *IFIP Working Conference on The Practice of Enterprise Modeling*. 2016. Springer.