Satu Data Indonesia in Sectoral Statistics: Concept of Satu Data Metadata Framework (SDMF)

H S Raja[1], **C A Adlan**[2]

[1]Statistics of Banjarbaru Municipality, South Kalimantan
[2]Gesellschaft für Internationale Zusammenarbeit (GIZ), Germany

*Corresponding author's e-mail: shandikaraja@bps.go.id

Abstract. Satu Data Indonesia is a policy contrive to encourage the problem of inadequate data governance in Indonesia. This policy makes 4 main principles, namely metadata, data standards, reference codes, and interoperability as metrics of success in its implementation. In this study, we analyze the satu data Indonesia implementation in Kutai Timur Regency. We found that the integration of the satu data principle is challenging to apply technically because sectoral data in Indonesia has 2 characteristics based on the preparation of the list of data needs, namely the centralized data list, and the decentralized data list. Decentralized data list is a list of data that is partially prepared in each agency without any coordination with other stakeholders for completing the satu data principles. To accommodating this condition, we design the Satu Data Metadata Framework (SDMF) a data standard framework that is in accordance with the conditions of data governance in Indonesia. SDMF utilizes contextual layer and discovery layer of metadata to provide temporal attribute called Satu Data Resource Identifier (SDRI) for integration purpose

1. Introduction

Data governance in Indonesia currently shows an improvement, one of which is the issuance of Presidential Regulation No. 39/2019 concerning Satu Data Indonesia (One Data Indonesia). This regulation provides direction for all institutions and communities that produce and manage data in Indonesia to apply the principles of open data that are adapted to the principle of one data Indonesia. There are 4 principles proposed in this regulation, namely one metadata, one data standard, one reference code, and the last is data interoperability. In addition, satu data Indonesia regulation also regulates the composition and access of organizations that can manage and run this policy.

The formulation of the Satu Data Indonesia policy is also followed by rules regarding the data planning process, managing data collection activities, and disseminating data. The purpose of managing the business process is to create quality and accountable data. From a technical point of view, this policy also provides guidelines for IT eligibility standards that can be used.

Before satu data Indonesia policy is published, there was a national statistical system regulated in the head of BPS regulation No. 5 of 2000. The national statistical system contains the flow of an implementation of sectoral, basic, and specific statistical data provision in Indonesia. The objectives of creating a National Statistical System, among others are to optimize the available resources used by the undertakers of statistical activities, to avoid duplication of statistical undertakings by the statistical undertakers, to create a reliable, effective, and efficient National Statistical System. However, in its implementation, the National Statistical System causes the endpoint of sectoral data to rely heavily on

BPS technical guidelines as a clearinghouse, this can be seen from the publication of sectoral data which is dominated by BPS output publications.

The role change induced by Indonesia's one data policy as a new component in the sectoral data supply business process in Indonesia has a direct influence on sectoral data provider agencies, including the vast number of stakeholders that engage in one data providing activity. The demand for human resources with strong IT abilities is high, but the supply is limited. This is due to the usual procedure of satisfying the principle of one data Indonesia, such as focus group discussions or other kind of formal discussion.

The aim of this paper is twofold. Firstly, we are implementing and analyzing the results of the implementation of the Satu Data Indonesia ecosystem in Kutai Timur Regency conducted by collaboration with Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ). The processes carried out include measuring the readiness of human resources, measuring infrastructure readiness, compiling regulations through regent regulations, and building a satu data portal (one data portal). After conducting this analysis, we developed a standard resource framework that can be used to facilitate the flow of Satu Data Indonesia implementation.

2. Implementation of Satu Data Indonesia in Kutai Timur Regency

In this section, we discuss the process of developing and executing the satu data Indonesia (one data Indonesia) policy in a regency in Kalimantan Timur, Kutai Timur Regency.

2.1. Study Area

We sorted Kutai Timur regency as the study area for this policy. This area was chosen in accordance with the following primary criteria: no government policy and implementation of a satu data policy.

2.2. Action Plan

Based on presidential regulation no. 39/2019, the implementation of satu data Indonesia is also carried out at the regional level, where the structure used is already regulated by this regulation. 4 roles must be filled, namely data coordinator, governing board, data custodian, and data producers.

From this regulation, we collectively with the relevant local government have compiled an organizational structure in to committees board for Satu Data Indonesia in Kutai Timur Regency, where the stakeholders included in the structure are Bappeda, Diskominfo, BPS, JIGD, and several agencies as data producers. Figure 1 shows the organizational structure of one Indonesian data in Kutai Timur.

Table 1. Satu Data Committees Board.

Role	Stakeholders
Coordinator	Planning and Development Board (Badan Perencanaan Daerah/Bappeda)
Data Custodian	Information and Communication Board (Dinas Komunikasi dan Informasi/Diskominfo)
Governing Board	Statistics Office (Badan Pusat Statistik/BPS) , Geospatial Information Network Agency (Jaringan Informasi Geospasial Daerah/JIGD)
Data Producers	Local and Sectoral Government Agencies

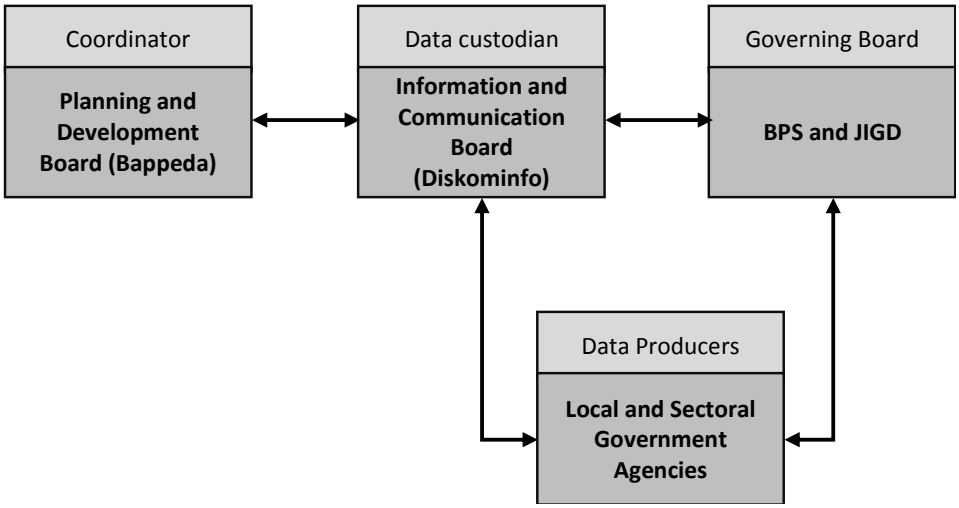

Figure 1. Organization Structure for Satu Data Indonesia in Kutai Timur Regency.

Other components prepared in the implementation of the Satu data Indonesia in Kutai Timur Regency include regulation in the form of a regent's regulation, this regulation is designed as a derivative of presidential regulation no 39/2019 so that the concept of satu data applied will be appropriate. Then from the technical side, BPS and Diskominfo designed standard operating procedures for data integration. Because of the short time to design, we prefer to use the existing framework and system, as we can see in table 2, we utilize CKAN as an open data framework and SIMDASI as a centralized data integration system owned by BPS. Next phase is we develop the guideline, then continued with human resource training related to the business processes of Satu Data. As we can see on table 3 shows the components of the action plan that was carried out for the implementation of one data set in Kutai Timur.

Table 2. Supporting System for Data Integration in Satu Data in Kutai Timur Regency.

System	Task	Operators
Comprehensive Archive Knowledge Network (CKAN)	Supporting Open Data Ability	Diskominfo
Sistem Manajemen Data Terintegrasi (SIMDASI) / Integrated Data Management System	Data Entry System for Priority Data List	BPS and Diskominfo

Table 3. Component of Action Plan Satu Data in Kutai Timur Regency.

Component	Output	Stakeholders
Regulations	• Regent Regulation	GIZ, BPS, Bappeda, Diskominfo
Information Technology Regulation and Requirement	• CKAN based Satu Data Portal • Simdasi guideline to data entry system	Diskominfo, BPS

Component	Output	Stakeholders
Training	• Application Programming Interface (API) training • Satu Data Business Process Training • Data Entry Training	GIZ, BPS, Diskominfo
End Point Product	• Satu Data Kutai Timur Portal	GIZ, Diskominfo

2.3. Implementation

Results of working on the components of the action plan for the implementation of satu Data Indonesia are following the expected output. The satu data (one data) regulation is stated in the regent's regulation no 49/2020. With this regulation, the application of the concept of satu data Indonesia in Kutai Timur can work properly. Concerning this regulation, the local government also held a discussion forum as the initiation of the Satu Data Kutai Timur Forum. The issue discussed is a list of priority data needs that will prioritize their availability in the data portal. The results of the discussion resulted in an agreement on the list of data that was used as priority data. Table 4 show the list of data priority.

Table 4. Priority Data List in Kutai Timur Regency.

Data	Source
Gross Domestic Regional Product (GDRP)	BPS
Poverty	BPS
Employment	BPS
Human Development Index (HDI)	BPS
Education	Education Board and BPS
Stunting	Health Board
Population	Civil Registration Board
Plantation	Plantation Board

Furthermore, in the human resource issue, GIZ provides training access to local government agencies related to the handling and development of API. Along with the action plans that have been designed, BPS, GIZ, and Diskominfo are trying to develop a satu data kutai timur portal, where the initial design used is to integrate SIMDASI and CKAN so that the data custodian and supervisor data functions can run in the same time.

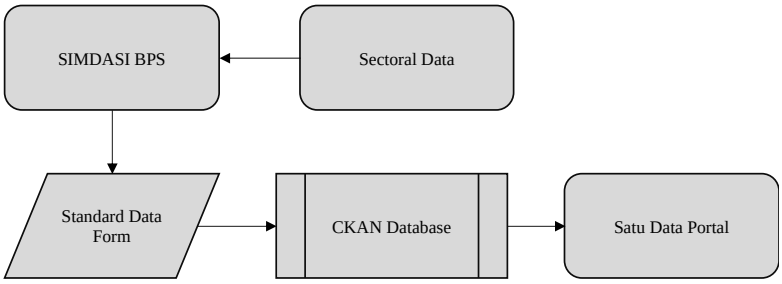

Figure 2. Satu Data Kutai Timur Portal flowchart.

In its implementation, portal development using the design shown in Figure 2 is quite difficult because the schema data and form of data are different between SIMDASI and CKAN. And finally, the portal building completely using CKAN.

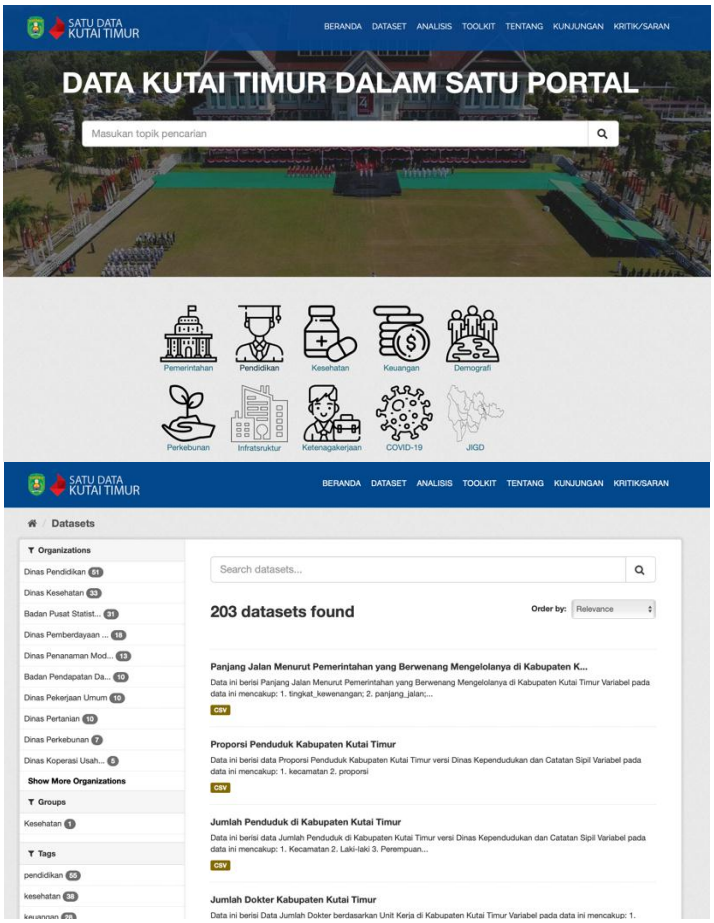

Figure 3. Satu Data Kutai Timur Portal Front Page.

2.4. Drawback Identified

As we can see, the process of implementing satu data Indonesia at the local government level requires a very huge effort in terms of planning for data provision. The long process of bureaucracy and expensive process for determining priority list data means that only a small part of the dataset will be focused on applying the satu data principle. As we see on table 5, only one data list that acquire from decentralized data.

Table 5. Classification on Priority Data List in Kutai Timur Regency.

Data	Source	Classification
Gross Domestic Regional Product (GDRP)	BPS	Centralized
Poverty	BPS	Centralized
Employment	BPS	Centralized
Human Development Index (HDI)	BPS	Centralized
Education	Education Board and BPS	Centralized
Stunting	Health Board	Centralized
Population	Civil Registration Board	Centralized
Plantation	Plantation Board	Decentralized

For data that have centralized classification, the completeness of the principle of satu data can be completed by data producers, because this data has its own procedure to measure its quality. For centralized data, the local government collaborates with BPS to jointly complete the one data principle, ensuring that the dataset's metadata and reference codes are met. For data that is not centralized and is not included in the priority data, it will be ignored. Based on field discoveries, this sort of data is solely used for evaluating and providing local or special information, and it is not included in the report to the vertical agency it covers. As a result, the planning process is directly dependent on the planner's understanding. If the planner understands the concept of Satu data, the resulting data will also follow these principles, and vice versa. This is illustrated in Figure 4. With the regional autonomy system in Indonesia, the amount of decentralized classification sectoral of data is quite large, so it is very necessary to build a solution to the application of the principle of satu data for all types of data regardless of priority and classification.

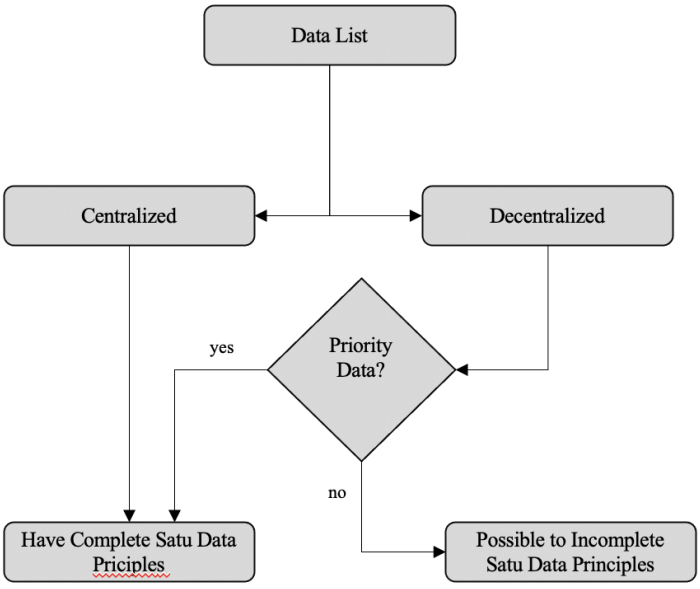

Figure 4. Effect of Using Data Priority List for Implementing Satu Data Principles

3. Optimizing Potential of Metadata

This section will deliver an overview of the current condition of metadata regulation in satu data Indonesia Policy and the advantages of metadata for integrating and completing principles of satu data during a partial business processes in different stakeholders is running.

3.1. Metadata concept in Satu Data Indonesia Policy

Metadata is generally used as the identity of data to describes various kinds of information related to the data. From the literature we used, there are 3 layers of metadata, namely discovery metadata, contextual metadata, and detailed metadata. The concept of these layers shown in Figure 5. The first layer is discovery (flat) metadata, this layer is very capable, simple to use, and allows to linkage several open datasets. The second layer is contextual metadata that allows providing more complete information from data, ranging from organization, publication, PIC, and various other aspects. And the last is the detailed metadata which is used to describe the specific dataset.

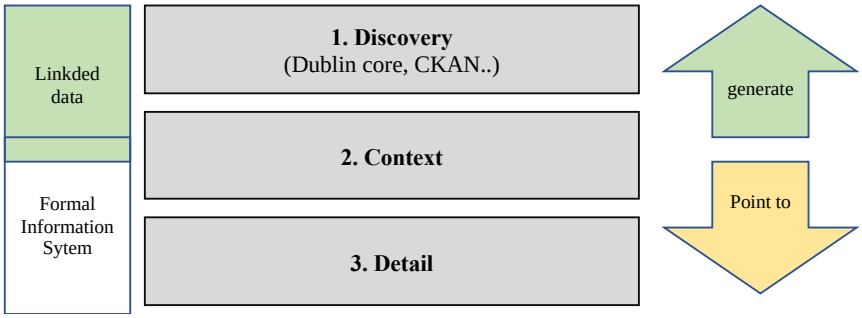

Figure 5. three-layers of metadata architecture.

At satu data Indonesia policy, there is no common regulation for metadata schemes that allow use in the universal datasets. The metadata used is still determined by the type of data and the governing board that handles it. In example, for sectoral statistical metadata using metadata concepts developed by BPS, metadata for financial datasets using metadata concepts from the ministry of finance. The type of metadata owned by this agency is contextual metadata but another aspect in the principle of satu data Indonesia there is an interoperability goal that requires a type of discovery (flat) metadata to be easy to access and connect between datasets. Based on these conditions, we can see that there has been a misleading in the use of metadata in satu data Indonesia policy.

3.2. Concept of Satu Data Metadata Framework

Base on resource, data in Indonesia is categorized into 3, there is sectoral data, basic data, and special data, this concept is established in the National Statistics System, in Figure 6 we can see the process of designing data requirements until the process of providing data is carried out. Each type of data is managed and produced by different stakeholders, basic data is produced by BPS, sectoral data is produced by sectoral governments, and special data is generated from various institutions or communities in the community. However, BPS was assigned the task of being a clearinghouse aimed at maintaining the quality of the data provided to the public.

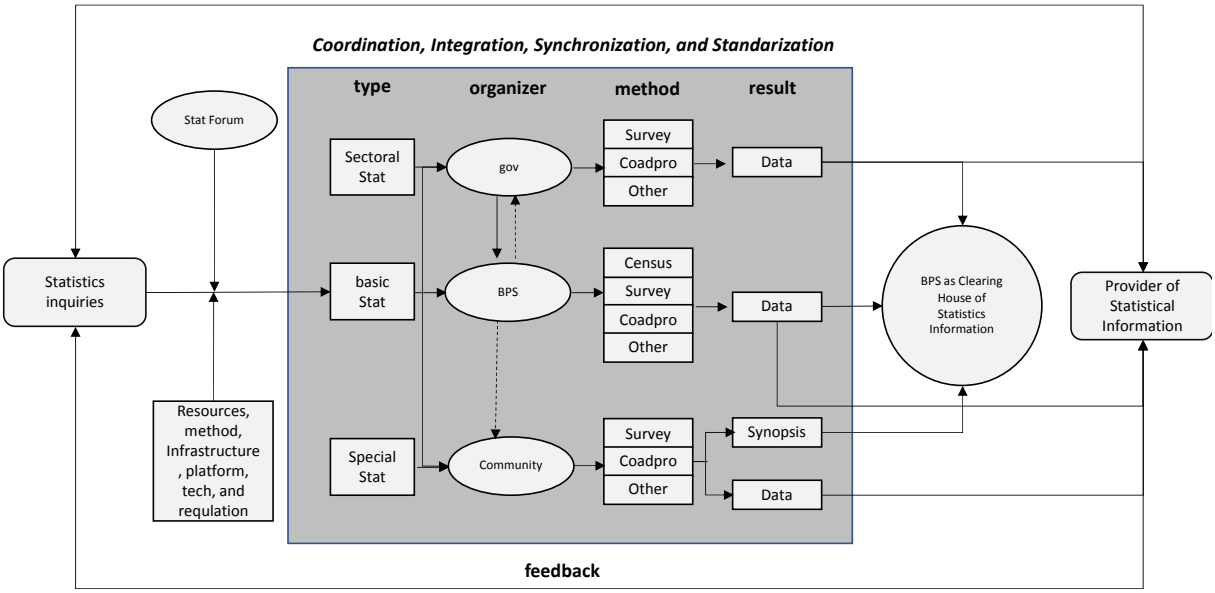

Figure 6. National Statistics System Architecture.

At a more specific level, the data production process, especially for sectors, is carried out partially according to needs and does not carry out coordination with other data producers. With the satu data Indonesia policy, the coordination scheme between agencies has been determined, in addition, the rights and obligations of each stakeholder have also been well described. With this regulation, the flow of data products from the lowest level planning to quality control of the resulting data is very clear, as are the actors implementing it.

This paper utilizes these conditions to develop satu data metadata framework (SDMF) that can explain and expose information about temporal relations in datasets, referential integrity, relationships between datasets, and the context of the datasets. This architecture utilizes the Statistical Metadata System (SMS) as a contextual metadata layer and Statistical Data and Metadata Exchange (SDMX) common vocabulary. Modification is done by adding an inheritance named satu data resource identifier in the IdentifiableArtefact class. SDMF can facilitate multi-stakeholder actors in the data production process, as shown in Figure 7.

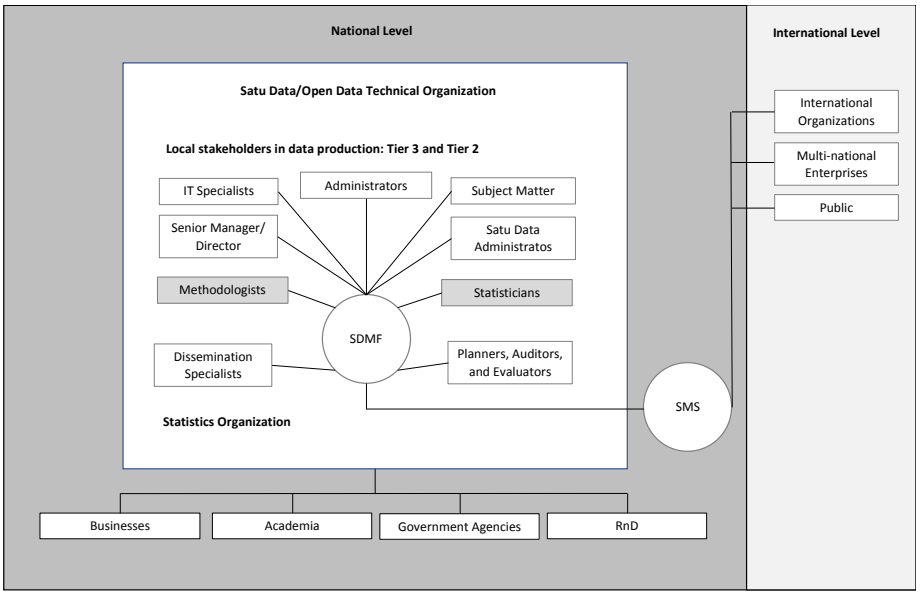

Figure 7. Actors of SDMF.

The architecture of the SDMF itself is a compilation of the utilization of discovery (flat) metadata using the Comprehensive Knowledge Archive Network (CKAN) and context metadata from the Statistical Metadata System (SMS). The two frameworks are integrated into the general SDMX schema, which will later become a metadata fulfillment vocabulary. Each stakeholder as a custodian and has access to add code to the relevant cell in SDRI. SDMF makes use of two classes from the SDMX schema: IdentifierArtefact for collecting SDRI information and Maintainer for managing information storage. The two classes are applied by including the parameters Sdri:int to capture the result of producing an integer SDRI code, followed by Ver:string to collect version control information from SDRI modifications and SetValue:string to keep the default value if paramater is null. Figure 8 shows the architecture.

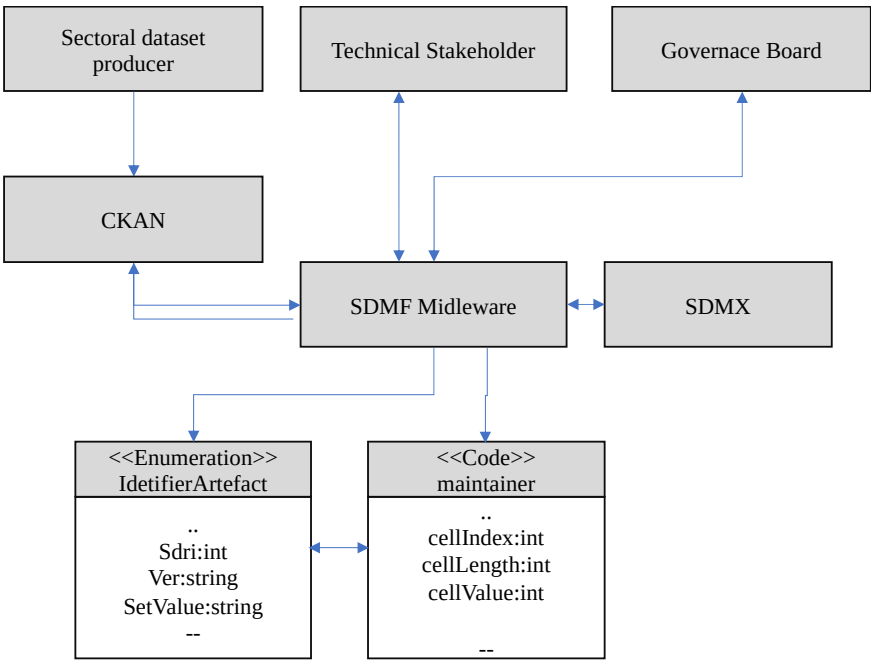

Figure 8. Architecture of SDMF.

3.3. Resource Identifier for Satu Data Metadata Framework

In this framework, there is one feature that is stored as a temporal identifier for each process that the data set goes through in compliance with the satu data principles. This feature is called satu data resource identifier (SDRI) in the form of an encapsulated serial number that can be filled in based on updates from each stakeholder who is responsible for determining the condition of the data set.

SDRI feature on SDMF allows metadata to provide a temporal relation of a dataset to its producer. Each cell will be filled if the dataset goes through the process to complete the principle of satu data Indonesia. concept of SDRI shown on figure 9.

SDRI Template

pp(2)	rr(2)	ic(3)	dl(1)	dc(3)	rc(2)	mc(2)	rd(6)	pl(1)	vc(3)
Example of Implemetation									
63	72	001	3	102	12	02	120221	0	002

pp	: province code	rc	: references code
rr	: reg code	mc	: metadata code
ic	: instantion/agency/stakeholder code	rd	: realese date
dl	: disemination levels code	pl	: priority list
dc	: data category	vc	: version control

Figure 9. Satu Data Resource Identifier (SDRI) template form.

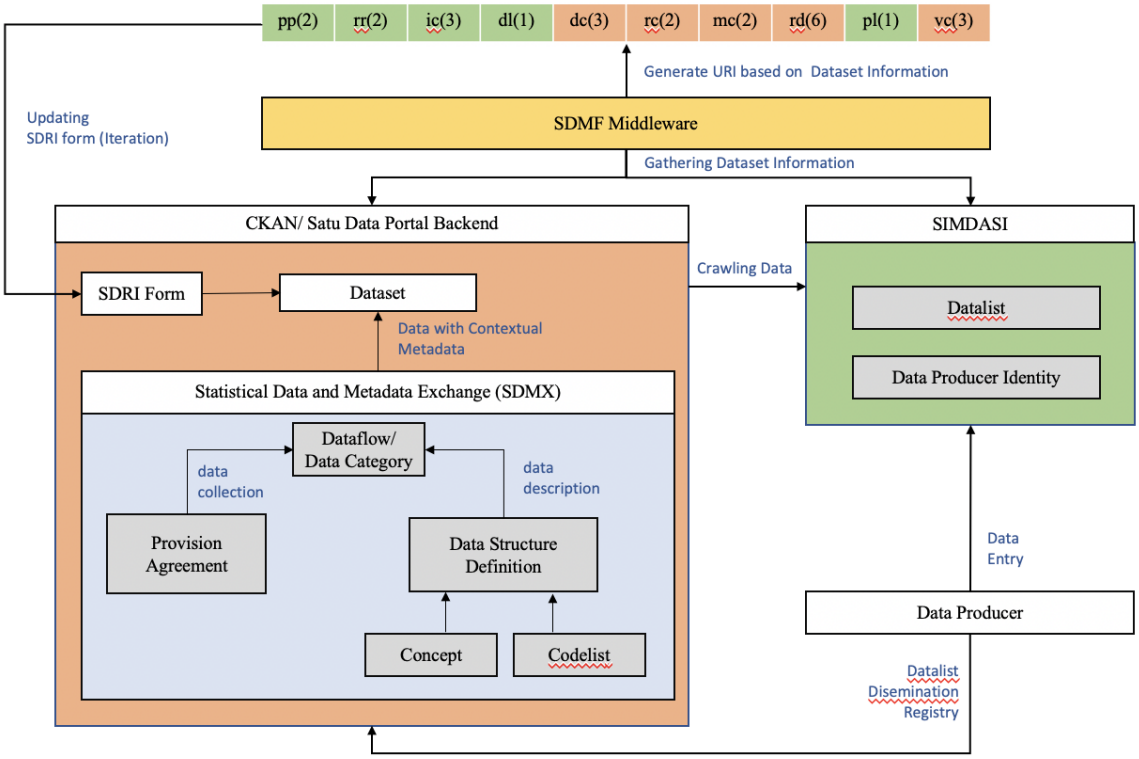

Figure 10. SDRI Updating Process.

The process for updating SDRI can be seen in Figure 10, the process starts from data producers inputting data in SIMDASI and completing the registration form for dissemination in the Satu Data Portal. Information related to the province code, regency code, agency code, dissemination level, and priority data is contained in the SIMDASI account owned by the data producer. While the version control information, data category, reference code, metadata code, and release date are contained in the dissemination form filled in by the data producer. The information is then collected by SDMF middleware to be converted into an SDRI form. Every time there is a change in version control, SDMF will retract information and update SDRI. The SDRI form is then automatically submitted to CKAN to complete the contextual metadata.

The Dissemination Registration Form can be completed on a regular basis, and data can be published even if the form is incomplete. The SDRI code provided in the metadata indicates the criterion of completeness of data information. The SDRI code allows the data governing board and

data coordinator to quickly conduct queries to automatically check data without having to go through manual requests from data producers. This will aid the process of monitoring decentralized and non-priority data in order to fulfill the satu data principle.

4. Conclusions

Satu data Indonesia policy is a positive response to the inadequate condition of data governance in Indonesia. This policy stipulates 4 principles that must be met to maintain data quality in Indonesia, namely one metadata, one reference code, one data standard, and data interoperability capabilities. To see the results of the policy, we implemented the concept of satu data in Kutai Timur Regency. The policy and all technical components for the implementation of satu data in Kutai Timur were successfully carried out, and the results showed several shortcomings, namely a large effort in coordinating the provision of data lists, decentralized sectoral data is lack of quality attention, the use of a single data portal which is limited to interoperability functions and does not accommodate the other 3 principles. the use of metadata that is not between the needs of integration and dissemination.

From these problems, we build the Satu Data Metadata Framework (SDMF) which is a modification of the Statistical Metadata System (SMS). This framework aims to accommodate the needs of ideal data integration and data dissemination in satu data policy. these needs are accommodated through SDMF's ability to store temporal resources and contextual descriptions of datasets. This capability is obtained from the addition of a new encapsulated serial number feature called Satu Data Resource Identifier (SDRI) in the IndentifierArtefact class in the SDMX common metadata vocabulary used.

Acknowledgement

The goal of this research is to create a satu data ecosystem in Kutai Timur. The authors would like to express their gratitude to those who assisted in the development of this work, particularly Diskominfo and Bappeda, as well as GIZ for assisting and participating in the implementation of Satu Data Indonesia in Kutai Timur.

References

[1] Matthews, B., Sufi, S., Flannery, D., Lerusse, L., Griffin, T., Gleaves, M., et al. 2010. *Using a core scientific metadata model in large-scale facilities*. The International Journal of Digital Curation. **5** 106-118.

[2] A Zuiderwijk., K Jeffery., M Jansse., 2012. The potential of metadata for linked open data and its value for users and publishers. *Journal of eDemocracy*. **5** p 222-244

[3] United Nations Economic Commission For Europe 2009. Statistical Metadata in Corporate Context. (Geneva). pp 14-15

[4] Zhang, J., Dawes, S. S., & Sarkis, J. 2005. *Exploring stakeholders' expectations of the benefits and barriers of e- government knowledge sharing*. Journal of Enterprise Information Management, **18**, pp 548-567.

[5] Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., & Sheikh_Alibaks, R. 2013. Socio-technical impediments of open data. Electronic Journal of eGovernment.Reference lists

[6] Javier D. Fernández, Miguel A. Martínez-Prieto, Claudio Gutiérrez et.all., 2013 *Binary RDF representation for publication and exchange (HDT)*, Journal of Web Semantics,Volume **19**, Pages 22-41

[7] Satu Data Indonesia National Secretariat, 2020. *Booklet Satu Data Indonesia* [online], Available: https://data.go.id/static-asset/unduhan/ [accessed august, 18 2021].