TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol. 14, No. 2, May 2015, pp. 343 ~ 352
DOI: 10.11591/telkomnika.v14i2.7505

Data Exchange Design with SDMX Format for Interoperability Statistical Data

Jaka Sembiring*[1], Ana Uluwiyah[2]
[1]School of Electrical Engineering and Informatics, Institut Teknologi Bandung,
Jl. Ganesha 10 Bandung 40132, Indonesia, Ph./Fax:+62-22-2502260/+62-22-2534222
[2]Education Training Center, Statistics Indonesia,
Jl. Jagakarsa Raya 70, Jakarta Selatan, Indonesia, Ph./Fax: +62-21-7873782-83/+62-21-7875497
*Corresponding author, e-mail: jaka@itb.ac.id

Abstract

Today's concept of Open Government Data (OGD) for openness, transparency and ease of access of data owned by government agencies becomes increasingly important. This initiative emerges from the demand of data usersforthe data belongs to the government agencies. The data services providing an easy access, cheap, fast, and interoperability are needed by the users and becomes important indicator performance for respective government agencies. Statistical Data and Metadata Exchange (SDMX) is a new standard format in the data dissemination activities particularly in the exchange of statistical data and metadata via Internet. In this respect SDMX support the implementation of OGD project. This paper is on the technical design, development and implementation of data and metadata exchange service of statistical data using SDMX format to support interoperability data through web services. Three results are proposed: (i) framework for standardization of structure of statistical publications data model with SDMX; (ii) design architecture of data sharing model; and (iii) web service implementation of data and metadata exchange service using Service Oriented Analysis and Design (SOAD) method. Implementation at Statistics Indonesia (BPS) is chosen as a case study to prove the design concept. It is shown through quantitative assessment and black box testing that the design achieves its objective.

Keywords*: open government data, statistical data and metadata exchange, web service*

1. Introduction

Initiative on Open Data Government (OGD) is increasingly getting supportfrommanycountries. Opendata provides an easy accessto the data so thatvarioususerssuch asinstitutions, scientistsorothercommunities are allowed to retrieve datawithouta licenseorpatentrestrictions [1]. Providing freely availabledatato everyonevia theWebiseasier when the data is interconnected each other. Opennessandconnectednessof datais the goalof theparadigm oflinkedopendata [2]. Open data becomes an important aspect in a public institution whose function is to serve the communities. With open data a public institution can fulfill its obligation of creating aprime serviceas one of their Key Performance Indicators for organizational performance [3]. Statistics Indonesia (BPS) as a public institution also has a strategic objective of delivering a prime service to the public in providing and disseminating high quality data and information. BPS is encouraged to change the method of disseminating data and information to be in line with the OGD initiatives. One of thecrucial parts is on how toexchange data and metadata that can provideinteroperabilityservices for openness and connectedness.Statistical Data and Metadata Exchange (SDMX), which is initiated by world seven statistical organizations, can be used as a framework in standardizing statistical data and metadata exchange. With SDMX it is possible to simplify and streamline the process of exchanging data and metadata usingthe same data structure and concept,in order to improve the timeliness of data services, accessibility, interpretability and coherence [4-6].In this paper we will use SDMX for standardizing the data and metadata where for clarity we will use BPS data as our special case.The SDMX standard will be implemented in Service Oriented Architecture (SOA) scheme described in [7], where in particular we will use the Service Oriented Analysis and Design (SOAD) as a design approach [8]. We provide the prototype of

Received February 11, 2015; Revised April 21, 2015; Accepted May 1, 2015

implementation of the design using web services. This is a natural choice since the web service is designed to support interoperability and interactions between system on a network [9]. In this paper we propose three main results: (i) standardization framework of statistical publication/aggregate data model structure, (ii) design architecture of data sharing service model, and (iii) implementation of design architecture of data exchangeusing web service. We will also present the validationresult of the proposed design with both user/expert judgment and black box testing.

2. Research Method

2.1. Standardization Framework of Statistical Data Structure with SDMX

This section will explain our first result on the standardization framework of statistical publication data structure. For case illustration we will use the business process of Data Dissemination Directorate (DDS) of BPS. The business process of the mentioned directorateis a process flow of activities to disseminate the result of statistical information in the form of services to the public users. The objective of our proposed scheme is to improve the value of public organization, in this case BPS, through increasing the openess of data, ease of access, and data interoperability. Openess of the data can be measured by availability of both data structure and content that can be accessed and used, reused and redistributed by all data users. Ease of access can be measured by data servicesmedium thatcan beaccessed easilythrough thedevicegadgets, mobile phones, PCandlaptop, etc. Meanwhile interoperability can be measured by capability service to be operated on different platforms.In order to achieve thesevalues and to show the conceptproposed in this paper, we adopt General Statistic Business Process Model (GSBPM) for business process flow [10]. By adopting GSBPM we can derive the business process as shown in Figure 1 whereas the business process of data disseminationcan be elaborated further in Figure 2. In this business process model, the data to be published will be taken from the dissemination data warehouse, not directly from subject matters. Based on the GSBPM, we propose the new business process architecture for data dissemination as shown in Figure 3. In this architecture, the published data is taken from a data warehouse after being validated and analyzed by Analysis and Development Statistics Directorate. Before publication, the data will be validated and mapped to the model structures/data scheme predetermined in the form of DSD/MSD. This process is conducted as a methodto standardize format and model structure of the published data, and to avoid the inconsistency of data. The data is ready to be published/released to the web service provider only after it is validated. We focus on the dissemination of statistics information through creating services of data and metadata exchange where for this purpose we createa new service for data and metadata exchange. This service provides easy access for the users.

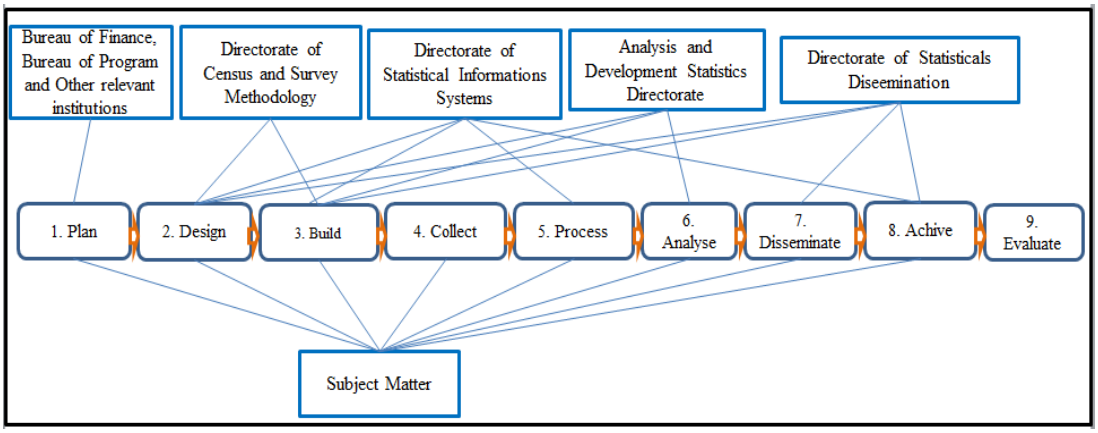

Figure 1. Proposed business process activities at Data Dissemination Directorate of Statistics Indonesia after adopting the GSBPM. Note that difference from existing condition, the data to be published will be taken from the dissemination data warehouse, not directly from subject matters

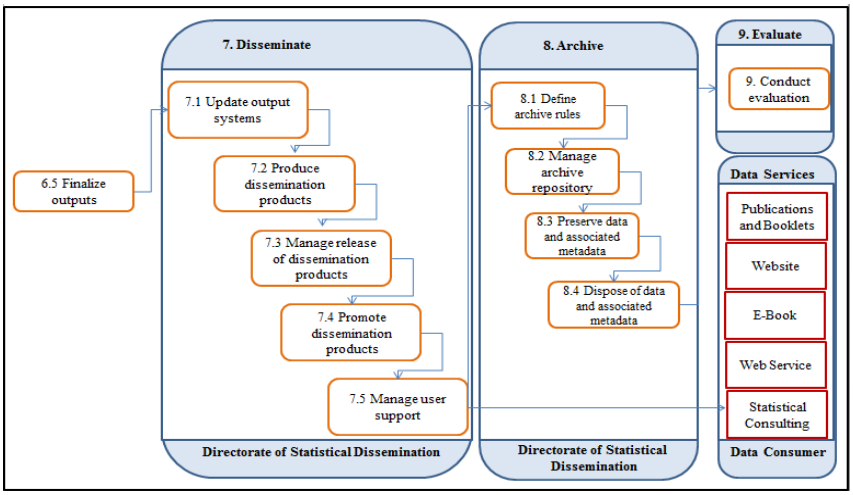

Figure 2. Detail business process of data dissemination derived from GSBPM. Note that this decomposition provides us with candidate services such as dissemination service

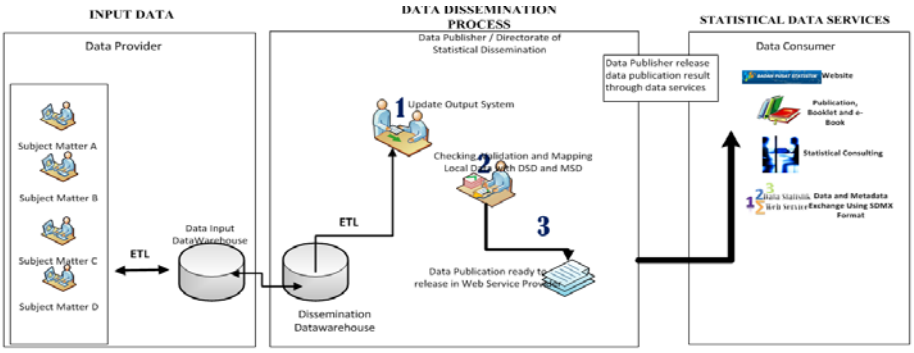

Figure 3. Proposed business process architecture of statistics data dissemination. This architecture has been aligned with the future contextual plan of BPS

The framework of data and metadata model structure is designed with SDMX-Information Model (SDMX-IM) methodwhich incudes several components: DSD, MSD, Data set, Metadata set and SDMX-ML message. Based on the analysis on the data to be published and SDMX-IM components, we can define the Level 1 standardization structure model as in Figure 4 where we have three main activities in this process as follows:

a) delivery of data to be published to the Data Dissemination Repository,
b) mapping local data with Data Structure Definition (DSD) and Metadata Structure Definition (MSD), and
c) publish/release of data publication.

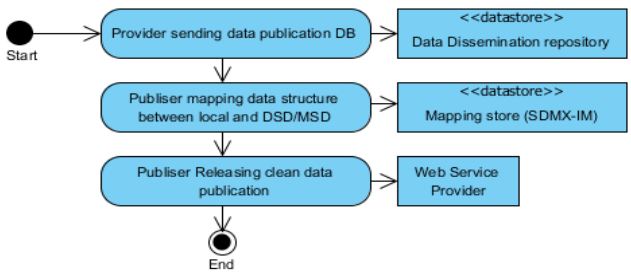

Figure 4. Standardization process data structure Level 1

We proceed to the Level 2 standardization of the model structure for detail process as in Figure 5.

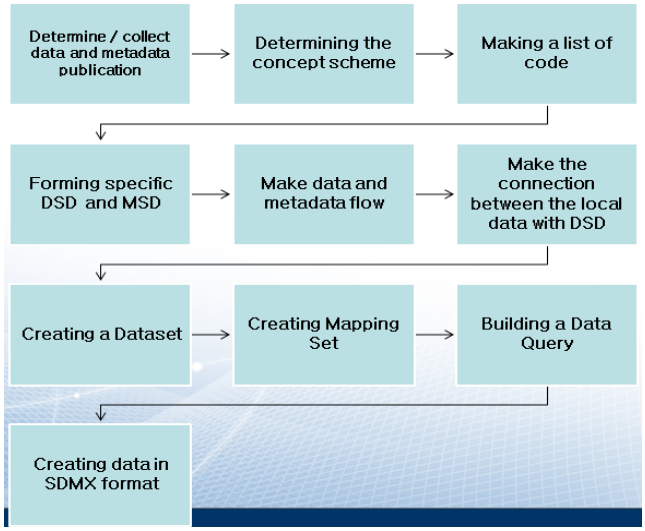

Figure 5. Standardization process data structure Level 2

The general process in Level 2 can be explained in detail as follows.

1) Determine the Concept and Concept Scheme. Concept has a very important role in the SDMX-IM because it is used to describe the multi-dimensional table structure or metadata structurere port. Concept structure consists of Dimensions, Measures, and Attributes. The sample result of concept and concept scheme in this paper can be seen in Table 3.
2) Creating Code List. Code list in the SDMX-IM is a list containing codes that represent concepts (dimension or attribute) in either DSD or MSD [11]. Sample of code list created in this paper is shown in Table 3.
3) Data Structure Definition (DSD). DSD is basically a description for concepts that have been identified and established. It describes whether the concept is a dimension or an attribute. DSD can be used in Time Series (TS), Cross Sectional (CS) and multidimensional data tables. In our case to create DSD, first we need to differentiate between TS data and CS data. Then we identify the concept, determine the measure and make an association with the Code Lists. The sample result for our case can be seen in Table 3.
4) Metadata Structure Definition (MSD). As in DSD, MSD is a structure for metadata. Metadata is used as a description or reference for the object to be exchanged. The purpose of metadata structure is to facilitate indexing, searching, processing and reporting of statistical activities. BPS devide metadata into three types: descriptive metadata, structural metadata and administrative metadata. These metadata are associated with reference metadata in SDMX.
5) Mapping Data. The main problem of data exchange on the data sharing systems is that there are various database structure integrated in one system, so that itis necessary to perform data mapping amongst related database to improve the convergence [12]. In our case we employ five steps for mapping between local databases and the predefined DSD concept as follows.
   a) List theconceptsthat existin thelocal data.
   b) Pairingalllocal dataconceptwiththe DSD concept.
   c) Pairinglocal codeswithDSDcodes.
   d) TranslatingSDMXQuery.
   e) Translate intodataset.
6) SDMX-ML message. SDMX-ML is created based on XML; so that SDMX complieswiththe concept and the rules of XML.The most important component in the implement SDMX XML are an XML namespace and XML Schema (XSD).

2.2. Architecture of Data Sharing Model

The general modelof dataandmetadataexchange processproposedin this paper isa simple model where dataproviderpublish/disseminatethe datatothe consumer using web services for data sharing implementation. The system diagram can be seen in Figure 6.

Figure 6. Data sharing model using web service

In our case, two models of datasharing are involved, i.e.: (i) datasharing from data provider to the publisher system, and (ii) data sharing from the publisher system to consumer. Data provider including subject matters and government institutions; sends the data to the system using push method, in other words the data provider push/send the data to the Data Dissemination publication repository. Whereas thedata transmission from system to consumers is using pull method where consumers ends a data request message and will pull/retrieve dataset fromweb service system. The architecture design of data dissemination that supports the above mentioned data sharing can be seen in Figure 7.

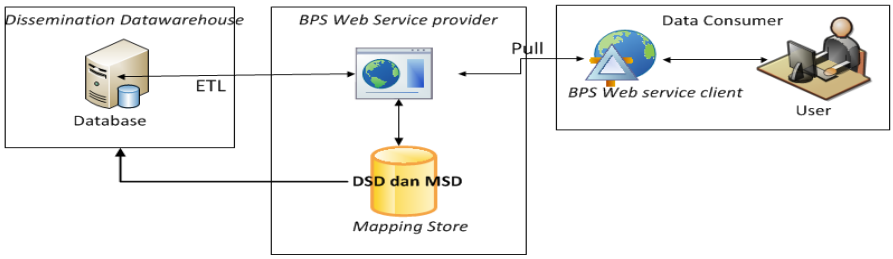

Figure 7. Detail of data sharing system design

The dissemination data warehouse saves all the local data publication comes from the subject matters and other data sources. Based on GSBPM, this warehouse will be managed by the Statisticals Information Systems Division and Statistical Dissemination Divisionso that the data can be validated before publication.In addition to dissemination repository, we have a mapping store to save the data concept. Statistical Dissemination Division who also performs the data mapping maintains the mapping store. With this architecture, all data will have the same concept. Both data storage are designed to support data exchange services. User is allowed to query or request the data via the web client, and this web client will send a message to the web service. This service will retrieve the data from dissemination data warehouse and mapping store located in the BPS servers.

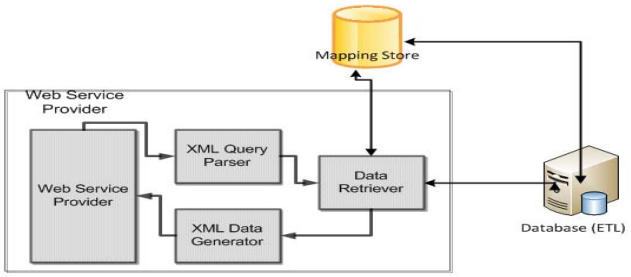

Figure 8. Data dissemination web service architecture

The web service architecture for data exchange services is shown in Figure 8. Users who send request data in SDMX Query messages, which will be translated by SDMX Query Parser, initiate the process of data dissemination in the web service. The SQLQuery retrieve the data structures and datasets from the store Mapping and data dissemination repository, and the system will generate data in SDMX-ML format which will be sent to the webclient.

2.2. Design of Data Exchange Service with SOAD

In this section we will discuss the design process of data and metadata exchange services. We use SOAD methodology as an approach of SOA [8]. Following the SOAD methodology for analysis and design, first step is the Conceptual View (CV) where we obtain the business process and sub business process for data and metadata service as shown in Figure 9.

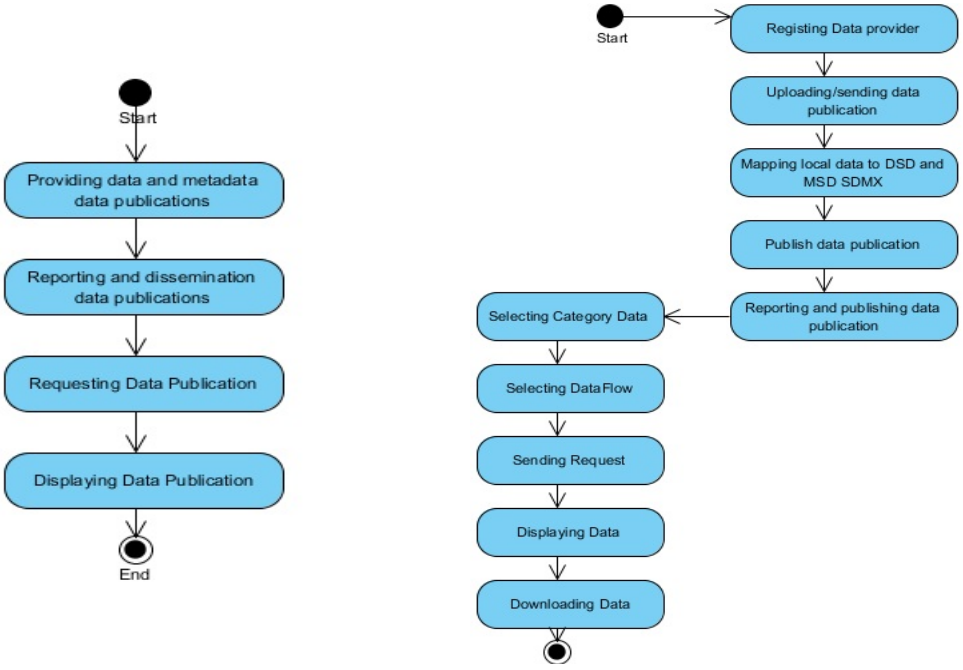

Figure 9. Business process and sub-business process analysis

Based the proposed system depicted in Figure 3, there are two systemswe need to create, i.e.(i) Web Service Provider, which is used by data publisher to provide, report and disseminate statistical data, and (ii) Web Application Client,which is usedbythe usersto invoke data.The general function of the system is described in Table 1.

Table 1. General function of the proposed system

Integrated Data Publication and Data Sharing Management System Service		Description
Web Service Provider	Providing publication data and metadata	Facilitatethe dataprovidertoprepareandtransmit data/metadata for the users
	Reporting and dissemination publicationdata	Facilitatetheinstrumentforpublication and reportingthe data for publicationsafterbeingvalidated
Web Application Client	Requesting publication data	Facilitatetheuserstorequestthe databased on the published categoryof data
	Displaying publication data	Facilitate the users by displayingdata fromthe userqueryaccording to the provided format

Second step, the Logical View (LG) consists of three layers: business layer, service layer and components layer which is derived from CV. The result of analysis can be seen in Table 2.

Table 2. Proposed services at Logical View

System FunctionalDomain	Coverageof Functional system	Description
Providing publication data and metadata	Registing data provider	Servicesfor registeringthe publication data
	Sending data publications	Servicestouploadthe publication datatothe database of dissemination/publication
	Mapping local data to DSD and MSD SDMX	Services used by datapublishertoverify, mapandvalidatethe publicationdatawiththe pre defined DSDandMSD
	Publishing the publications data	Servicesused by dataprovidersto approvethe publication datato be published to the public
Reporting and dissemination publication data	Reporting and dissemination publication data	Servicefor the users to view the published data
Requesting publication data	Selecting data category	Servicesfor the usersto selectappropriatedata categories
	Selecting data flow	Servicesfor theusersto choosedataflow
	Sending request	Servicesforthe userto choose andinvokethe datarequest
Displaying publication data	Displaying data	Servicefordisplaying datawithSDMX-ML, PDFandXLS data format,andfor visualizationwithtabulationandgraphs
	Dowloading data	Service for downloading datawith specificformat

Finally, the Physical View (PV) consists of four layers web service layer, presentation layer, appplication layer and data model layer. Web service layer is divided into two parts: web service design and service sequence diagram. Presentation layer is an interface design for Integrated Data Publication and Data Sharing Management System where in this paper we provide seven interfaces including: Home Page, Data Category, Data and Metadata Set Upload, Data Mapping, Login/Register, Web Service Data Published, and Dataset Presentation. Application Service layer is the interfaces for business logic implementation constist of eight interfaces including: IdataCategory AppS, IdataMetadataSetUploadAppS, IdataMappingAppS, IloginAppS, ImemberRegisterAppS, IdataFlowAppS, IdataPresentationAppS, and IDSDMSDAppS. Data model layer consists of two Data Transfer Object (DTO). The first is the Data Dissemination DTO consists of LogQuery, DataSet, MetadataSet, Log Download and Publisher. This componentis used to store local publications data i.e. dataset and metadataset. The secondis Mapping Store consists of DSD, MSD, Mapping, Master Data Provider, CodeList, Category Data and Data Flow. This component is used to store data check, validation and mapping so that the data is readyto be disseminated.

3. Results and Analysis

In this section we will show the implementation result of the design involving public domain data on population/demography of Statistics Indonesia. Aggregate data for sample case is(i) the population of Indonesia by province 1971, 1980, 1990, 1995, 2000, and 2010, and (ii) total and percentage of poverty and poverty line by province 2007, 2008, 2009, 2010, 2011, 2012. Both data can be viewed in two categories, which are Time Series (TS) and Cross Sectional (CS). Using the method described in the previous section, first we determine the concept where in this simple case is denoted with POP_CONCEPT since we deal with data on population/demographic. Based on the concept, we can derive the code list such as in column five of Table 3. We can create the DSD/MSD by first separating the TS and CS, then using the method described in the previous section we can obtain the DSD/MSD where the sample result can be seen in Table 3.

Table 3. Format of Message Implementation Guide (MIG) DSD

DIMENSIONS				
Position in key	Concept Scheme			Codelist
	ID	Name	ID Scheme	ID CodeList
1	FREQ	Frequency	POP_CONCEPTS	CL_FREQ
2	PROVINCE	Population country	POP_CONCEPTS	CL_PROVINCE
3	INDIC_POP	Population Indicator	POP_CONCEPTS	CL_POP_INDICAT
4	POOR_POP	POORNESS Measure	POP_CONCEPTS	CL_POP_POOR
TIME	TIME_PERIOD	Time Period	POP_CONCEPTS	
MEASURES				
TYPE	Concept Scheme			Codelist
	ID	Name	ID Scheme	ID CodeList
Primary	OBS_VALUE	Observation value	POP_CONCEPTS	
CS	ID-AC	Aceh		CL_PROVINCE
	ID-BA	Bali		
	ID-BB	Bangka-Belitung		
	ID-BE	Bengkulu		
ATTRIBUTES				
Attachment Level	Concept Scheme			Codelist
	ID	Name	ID Scheme	ID CodeList
Observation	OBS_STATUS	Status of the observation	POP_CONCEPTS	CL_OBS_STATUS
Series	UNIT	Unit	POP_CONCEPTS	CL_UNIT
Series	TIME_FORMAT	Time Format	POP_CONCEPTS	CL_TIME_FORMAT

We continue presenting the result with the implementation of webservice provider which can be seen in Figure 10.

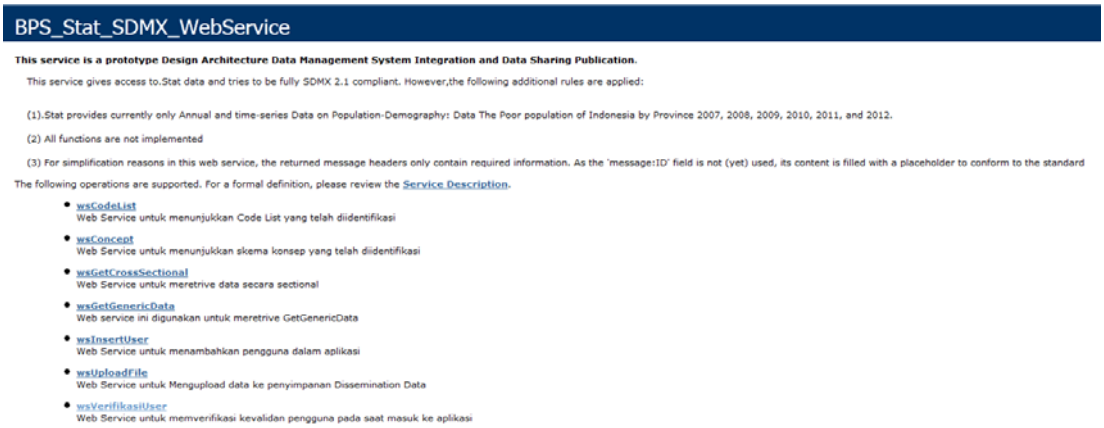

Figure 10. Implementation example of web service provider

We verify the standardization and architecture design quantitatively using simple Likert scale by presenting the result to the user/expert, which in this case is the relevant personnel at Directorate of Statistical Dissemination, Statistics Indonesia (BPS).Forthe web service implementation we perform both quantitative evaluation through Likert scale and black box testing using soapUI software. We obtained that for quantitative assessment the result is 22.3 in

Likert scale, whichmeans*very good*. Meanwhile for black box testing resultsusing soapUI can be seen in Figure 11. We conclude that all of the designed service perform as intended.

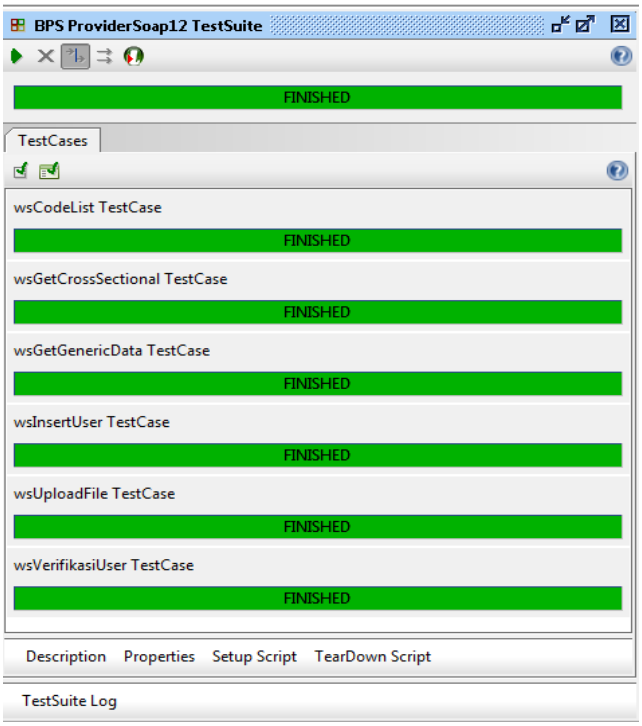

Figure 11. Result of black box testing using soapUI

4. Conclusion

In this paper we report on the design of data and meta data exchange using SDMX format for statistical data at Statistics Indonesia (BPS). We implement the design concept with web services. Implementation at Statistics Indonesia (BPS) is chosen as a case study to prove the design concept. We describe three main results. The first result is on the framework for standardization of structure of statistical publications data model with SDMX. The simple case we presented has been able to demonstrate the advantageous of the effort for interoperability of data and metadata. Second result is on the design architecture of data sharing model. The design is based on the user needs where we able to align the design with the long term master plan of BPS. Finally, the third result is onthe implementation of data and metadata exchange service using Service Oriented Analysis and Design (SOAD) method. We present the prototype system implementation with web services.It is shown through quantitative Likert assessment and black box testing that the overall standard, architecture design and implementation have achievedtheirrespective objectives.

References

[1] Hoxha J, Brahaj A. *Open Goverment Data on the Web: A Semantic Approach*. International Conference on Emerging Intelligent Data and Web Technologies. Tirana, Albania. 2011: 107-113.
[2] Bizer C, Health T, Berners Lee T. Linked Data-The Story So Far. *International Journal on Semantic Web and Information Systems (IJSWIS)*. 2009; 5(3): 1-22.
[3] Zakaria Z. Key Performance Indicators (KPIs) in the Public Sector: A Study in Malaysia. *Canadian Center of Science and Education*. 2011; 7: 102-107.
[4] Statistical Data and Metadata eXchange. 2012. sdmx.org/page_id=10.
[5] SDMX Standards Section 1 -Framework. 2013. www.sdmx.org.
[6] SDMX Standards Section 1 -SDMX-Information Model. 2013. www.sdmx.org.
[7] Erl T. Service Oriented Architecture: Concept, Tehnology and Design. New York: Prentice Hall PTR. 2005.

[8] Sarno R, Herdiyanti A. A Service Portfolio for an Enterprise Resource Planning. *International Journal of Computer Science and Network Security*. 2010: 144-156.
[9] Web Services Architecture. 2004. http://www.w3.org/TR/ws-arch/.
[10] Vale S. GSBPM. http://www1.unece.org.
[11] Cross-Domain Code List. 2009. www.sdmx.org.
[12] Ning L, Bin X, Xin Z. *Database Conversion Based on Relationship Schema Mapping*. International Conference on Internet Technology and Applications (iTAP). Wuhan. 2011: 16-18.