Development of big data system of LP3 UNNES to support excellent service of LPTK on international reputation

Sugianto\textsuperscript{1,2*}, Ngabiyanto\textsuperscript{1,3}, A Prabowo\textsuperscript{1,2}, A Arfriad\textsuperscript{1,4}

\textsuperscript{1}Board of Education and Profession Development (LP3) UNNES
\textsuperscript{2}Mathematics and Science Faculty, UNNES
\textsuperscript{3}Social Science Faculty, UNNES
\textsuperscript{4}Engineering Faculty, UNNES

Corresponding Author: sugianto@mail.unnes.ac.id

Abstract. Semarang State University, Education and Professional Development Institute (LP3) is a public service institution integrated with the Semarang State University Institute (UNNES). As a public service institution, LP3 is bound by regulations related to service satisfaction and service statistics as part of public accountability. One of the things that influence public accountability for the services of government institutions is the availability of data that can be accessed by the public [2], [3]. Apart from being used as a public service data archive, the data can also be used by the community as a system that supports decision making [4]. The system is then called the Decision support system (DSS).

Decision support systems or DSS are used to collect data, analyze and shape collected data, and make correct decisions or build strategies from analysis, no influence on computers, databases or human users [5]. Information that is usually collected using a decision support application will do: Access all the latest information assets, including legacy and relational data, data compilation, data warehouses, and large amounts of data collection; Sales figures from one period to another; Estimated income figures, based on the assumption of new product sales; The consequences of different decision-making choices, with experience in a context that is detailed again [6].

There are already many companies in various industries that rely on decision support tools, techniques and modeling to help them analyze and solve a variety of everyday business questions. Decision support systems are dependent on data, as the whole process takes all available data sets, for
analysis. Business Intelligence-based reporting tools, processes and methodologies are examples of important uses in any decision support system, and provide users with reliable data analysis, reporting and monitoring [7].

The usual requirements for implementing a High Level Decision Support System include: Collecting data from various sources (sales data, inventory data, supplier data, market research data, and other supporting data); Formatting and data usage; and Location of suitable databases and development of formats for report making and analysis based on decision making; and Devices and applications that are versatile and able to provide reporting, monitoring and analysis of data.

In fact, as a public service institution overseeing services in the fields of PPL, PPG, MKU / MKDK, Career and Counseling, as well as Language and Training services, LP3 does not yet have a qualified DSS. Broadly speaking, LP3 which is part of the Public Service Agency (BLU) University of Semarang State University, until 2018, LP3 has managed the number of active PPG Semarang State University (UNNES) students of approximately 1240 students, and was planned for 2019 will receive as many as 2000 PPG students. In addition to managing PPG students, LP3 also manages PPL programs domestically and abroad for Semarang State University students. LP3 also manages the certification process for Teachers in Central Java and Lecturers in Semarang State University. With so many activities carried out by LP3, it raises issues related to big data that must be managed in order to realize data validity to support excellent services of LPTK with an international reputation. The process of managing big data certainly requires a system that can integrate large amounts of data and can be monitored in real time which can function as a supporter of LPTK prime services internationally.

LP3 services that can be utilized based on the DSS include: (1) Identification of lecturers who have / have not been certified at UNNES; (2) Identification of lecturers who have / have not followed AA or PEKERTI; (3) Identification of regions or cities of origin of PPG participating teachers that are suitable for use as a target location for community service and research activities; (4) Determination of academic policies related to English language skills; (5) Determination of the location of schools where regular PPL and PPL PPG activities; Predictions of PPG graduation in the current year; and other necessary requirements.

These various data are arranged in a big data on the decision support system to improve LP3 accountability as a public service institution. In the end, a study on the development of UNNES LP3 big data system to support excellent service of LPTK with an international reputation needs to be done.

2. Theoretical Basis

2.1. BigData Technology

The rapid growth of internet technology has led to a variety of major activities at tertiary institutions using information technology. This results in huge data that is called big data about education and teaching. All these activities from time to time cause digitization of data.

The concept of big data can be interpreted as a collection of large data that is very large and very fast-growing that comes in various forms and has a certain value. Big Data technology has the ability to handle a variety of data variations. In general, there are 2 groups of data that must be managed, namely: Structured data, namely groups of data that have defined data types, formats and structures. The data source can be in the form of transactional data, OLAP data, traditional RDBMS, CSV files, spreadsheets; and Unstructured data, namely textual data groups with erratic format or no inherent structure, so to make them structured data requires more effort, tools, and time. This data is generated by internet applications, such as log URL data, social media, e-mail, blogs, videos, and audio [8].

The big data management stages include: Acquired, related to the source and how to get it; Accessed, related to data access power; data already collected requires governance, integration, storage and computing so that it can be managed for the next stage; Analytic, related to the information to be obtained, the results of data management that has been processed. The analytics can be in the form of descriptive (data drawing), diagnostic (looking for cause and effect based on data),
predictive (predicting future events) and prescriptive analytics (recommending choices and implications of each option); and **Application**, related to the visualization and reporting of results from analytics [9].

Technologies used in big data include:

**No SQL Store.** No SQL Store provides scalable data storage with a more relaxed consistency guarantee compared to traditional RDBMS. No SQL Store also provides a flexible scheme to allow heterogeneous columns on different rows of storage. [10]

**Real Time Query over HDFS.** Although the Map Reduce framework was very scalable for Big Data requests, it usually did not provide the real-time response that interactive queries require. Some solutions such as Impala try to solve problems using the real-time ad-hoc SQL query processing engine directly over HDFS, bypassing Map Reduce processing to allow for shorter response times. Additional optimizations such as compression can be used to speed up response time further. Solutions in this category provide horizontal scalable implementations for real-time requests at the Big Data level [10].

**Reduce and Hadoop map.** In this technology, cluster computers were used that work in parallel to process very large volumes of data. Then the data flow was separated into two phases namely the map phase and the reduction (reduction) phase. In the map phase, chunks of data were processed in isolation by tasks called Maps. The output from this map maker was brought to tasks called reduction, which produces the final output. Hadoop was an implementation of the Map Reduce framework. Storage using The Hadoop Distributed File System (HDFS) provides storage support for the Hadoop Framework. HBase provides additional distributed database functions via HDFS. Data stored in HDFS was processed by the MapReduce operation [11].

![Figure 1. Map Reduce Framework dan Hadoop](image_url)

### 2.2. Decision Support System of LP3 UNNES

Decision Support System or Decision Making System could be defined as a computer-based information system that serves to support decision-making activities in an organization [12]. This DSS was an interactive software-based system designed to help policymakers gather or compile various information that is useful for solving a problem and making a decision or solution to the problem. In this context, it could be simply defined as a computer-based information system that was related to various information about LP3 services and data related centers in LP3 UNNES. It could be used by decision-makers to find out the latest information and determine the direction of management in the coming years.

Understanding the importance of the information system, LP3 UNNES has initiated DSS by collecting data in separate files in the form of Excel. The file was stored on a drive that was stored on the network which could be downloaded at any time. The data was collected in monitoring activities carried out by the quality assurance group every 3 months. However, the use of online documents was not flexible. Not everyone could easily apply the file online. These weaknesses could be covered by a robust system for providing data.
In a simple workflow preparation and management of DSS can be described below. The technical implementation team as well as DSS management is under the LP3 Quality Assurance Group, and in its implementation is supported by the LP3 Secretary and Subbag in LP3. In accordance with agency policies, the LP3 web data system that has been built is included in the UNNES LP3 website.

![Figure 2. Workflow of LP3 UNNES’s DSS](image)

3. Method
Broadly speaking the research method used to achieve the goal was divided into 3 main parts, as shown in Figure 3.

![Figure 3. An outline of the research method carried out](image)

An outline of the research method used in detail was explained as follows: **Retrieve Data** was the process of collecting UNNES LP3 data available in the last 5 years. This process was realized in the form of raw data which will then be divided into several sections according to the needs of big data analysis. The data was stored in a centralized database; **Development System**. The system development process was divided into 4 parts, namely: **Requirement Analysis; System Design; Implementation; Integration and testing; Operation**. This final stage aims to run a system that was directly used by the user.

3.1. Kind of Data
Data collected in this study were grouped into two parts, namely: (1) data from the development stage in the form of data from the review of software engineering experts, data from laboratory / vocational expert review results, and data from review results from product users, namely laboratory assistants, (2) data from the evaluation stage in the form of data from product trial results to laboratory performance in the field. All data obtained are grouped according to their nature into two namely qualitative data and quantitative data. Qualitative data obtained from the development stage and quantitative data obtained from the evaluation stage.

3.2. Data Collection Instruments and Techniques
Data collection instruments used in this study were in the form of: (1) a validation questionnaire and (2) an observation sheet in the form of a performance assessment instrument in the form of a numerical rating scale.
3.3. Data analysis technique
This data analysis technique is carried out to infer about existing data. In this study data obtained through interviews at the potential problem stage are presented descriptively and interpreted qualitatively. In the pilot phase, the data is used to test the research hypotheses. In experimental research designs, there are usually correlated samples, so the hypothesis used is a type of comparative hypothesis [13].

4. Result and Discussion

4.1. Research Result
The result on retrieve data process show us that data on LP3 should be managed well. There was plenty of data that was not on LP3 Databases, such as: the mark of PPG student, registrant of PPG, and selection process of PPG. Those data were managed by national government. LP3 was just accepted data from BELMAWA and GTK.

The result on Development System could be accessed on data.lp3.unnes.ac.id. There were some parts about interface, data display, and manipulative data display. Interface was part of DSS which was related to user. Data display was related to main and raw data. The system just shows the data from the database of LP3 UNNES. The main and raw data such as: the number of PPL student on year XXXX, the number of UNNES TOEFL Test on period XX year XXXX, the number of PPG Student, the number of PPG Final test (UKMPPG), the registrant of International PPL Student, and another data from the center of LP3. The manipulative data display shows us the analysis of raw data. Examples of manipulative data display were the prediction of PPL International Registrant, percentage of PPG student who pass the test at the time, and prediction of the number of UNNES TOEFL on year XXXX. The manipulative data display could be used to prepare about thing related.

The example of the system showed on the figure 4 and figure 5 below.

![Figure 4. Interface of DSS LP3 UNNES.](image1)

![Figure 5. The Display of Raw Data of the number of PPG Student](image2)

4.2. Discussion
Based on the results and discussion, it can be seen that building an LPS DSS requires quite a long time. The data in LP3 comes from inhomogeneous forms and varied sources. It means the DSS was being prepared for a sustainable system [13], [14]. The database currently being built only strengthens the system at PPG and several other LP3 centers, such as Careers and Counseling, International PPL, and language services. The MKU Center will need more time to develop a database for DSS. The complexity of the data that builds the DSS is also conveyed in previous studies that big data will collect data from various sources [15], [16]. However, if the research is completed, this DSS will greatly assist the stakeholder in deciding policy. This result strengthening the previous research about DSS [17], [18], [19].
5. Conclusion

Based on results and discussion, we conclude that the DSS of LP3 UNNES was going well. The website of data.lp3.unnes.ac.id can be utilized by the community as a decision support system.

References
[1] Peraturan Menteri PAN dan RB No. 15 Tahun 2014
[2] Kurniawan T 1988. *J. Ilmu Adm. Dan Organ.*
[3] Nugraheni and Khaerunisa L 2016. *J. Ilm. Univ. Bakrie*
[4] Irvin R A and Stansbury J 2004 *Public adm. rev.* **64** 1.
[5] Aronson J E, Liang T P, and Turban E 2005 *Decision support systems and intelligent systems* Vol. 4 (New York: Pearson Prentice-Hall).
[6] Edwards J S 1996 *J. Oper. Res. Soc.* **47** 5.
[7] Kaklauskas A 2015 *Biom. Intell. Decis. Support* (Springer, Cham) pp. 31-85.
[8] Michalik P, Štofa J, and Zolotova I 2014 *Proc. Int. Symp. Appl. Mach. Intell. Inform.* p. 331.
[9] Gandomi A, and Haider M 2015 *Int. J. Inf. Manag.* **35** 2.
[10] Hu H, Wen Y, Chua T S, and Li X 2014 *IEEE access.* **2**.
[11] Almeir M H 2012 J. Emerg. Trends Comput. Inf. Sci. **3** 4.
[12] Marakas G M 2003 *Decision support systems in the 21st century* Vol. 134. (Upper Saddle River, NJ: Prentice Hall).
[13] Aiello G, Giovino I, Vallone M, Catania P, and Argento A 2018 *J. Clean. Prod.* **172**.
[14] Kukar M, Vračar P, Košir D, Pevec D, and Bosnič Z 2019 *Comput. Electron. Agric.* **161**.
[15] Sugiyono 2011 *Metode Penelitian Kuantitatif, kualitatif dan R & D* (Bandung: Alfabeta) p. 90.
[16] Barnett G O, Cimino J J, Hupp J A, and Hoffer E P 1987 *Jama.* **258** 1.
[17] Sim L L W, Ban K H K, Tan T W, Sethi S K, and Loh T P 2017 *PLoS One.* **12** 2.
[18] Dweiri F, Kumar S, Khan S A, and Jain V 2016 *Expert Syst. Appl.* **62**.
[19] Noorollahi Y, Yousefi H, dan Mohammadi M 2016 *Sustain. Energy Technol. Assess.* **13**.