KnE Life Sciences

Knowledge E
Engaging minds

Conference Paper

SIDeKa: The Role of Information Technology for Knowledge Creation

Cungki Kusdarjito

Faculty of Agriculture, Janabadra University, Jl. Tentara Rakyat Mataram 55-57, Yogyakarta 55231, Indonesia

Abstract

The advancement of big data analytics is paving the way for knowledge creation based on very huge and unstructured data. Currently, information is scattered and growth tremendously, containing many information but difficult to be interpreted. Consequently, traditional approaches are no longer suitable for unstructured data but very rich in information. This situation is different from the role of previous information technology in which information is based on structured data, stored in the local storage, and in more advanced form, information can be retrieved through internet. Meanwhile, in Indonesia data are collected by many institutions with different measurement standard. The nature of the data collection is top-down, carried out by survey which is expensive yet unreliable and stored exclusively by respective institution. SIDeKa (Sistem Informasi Desa dan Kawasan/Village and Regional Information System), which are connected nationally, is proposed as a system of data collection in the village level and prepared by local people. Using SIDeKa, data reliability and readiness can be improved at the local level. The goals of the SIDeKa is not only local people have information in their hand such as poverty level, production, commodity price, the area of cultivated land, and the outbreak of diseases in their village, but also they have information from the neighboring villages or event at the national level. For government, data reliability will improve the policy effectiveness. This paper discusses the implementation and role of SIDeKa for knowledge creation in the village level, especially for the agricultural activities which has been initiated in 2015.

Keywords: big data analytics; SIDeKa; unstructured data.

Corresponding Author:
Cungki Kusdarjito
ckusdarjito@janabadra.ac.id

🔓 **OPEN ACCESS**

1. Introduction

Data availability and its reliability in Indonesia can be considered as luxury good. Andrinof A. Chaniago, former Indonesian Minister of Development Planning admitted during his opening remarks on International Conference on Data Innovation for Policy Markers, carried out in Bali, 26-27 November 2014, that he had data thief due to overprotective officials and closed data systems had often pushed him to look for hidden

data for his own policy research work. In Indonesia, data are scattered among different institutions, yet for the same data. They are collected separately and repeatedly for the same data. For the village officers, entering multiple times for the same data requested by different institutions is a tedious work. As a result, data validity could be questioned because the village officer does not always fill the actual data due to deadlines and fatigue. Sometimes, these works are handed over to other person, such as the internet kiosk operators.

With the advancement of the information technology, new methods of data collection and analysis require more demanding statistics than ever before. This includes looking for ways to supplement official statistics. Although official statistics are generally the gold standard for data, their reliability overtime are in doubt due to inconsistency for data categorization. Moreover, official data when becomes available is out of date (lagging) so it is become constraint when trying to understand the phenomena that change quickly. For instance, the number of poor people who eligible for assistantships cannot be determined correctly which will lead to social tension, or even worse conflict, among the people. This situation often makes local leader in uneasy situation. In addition, at the macro level, decision makers do not have enough valid data and the related cost may increase.

As stated in OECD book [1], data quality should follow seven dimensions. First, data is relevant, characterized by the degree to which the data serves to address the purposes for which they are sought by users. Second, data accuracy represents the degree to which the data correctly estimate or describe the quantities or characteristics they are designed to measure. Third, data credibility, refers to the confidence that users place in those products based simply on their image of the data producer, i.e. the brand image. Confidence by users is built over time. One important aspect is trust in the objectivity of the data. Fourth, data timeliness reflects the length of time between their availability and the event or phenomenon they describe. Real-time data is data with a minimal timeliness. Fifth, data accessibility reflects how readily the data can be located and accessed. Sixth, data interpretability, reflects the ease with which the user may understand and properly use and analyze the data. Seventh, data coherence reflects the degree to which they are logically connected and mutually consistent. No matter how much structured and unstructured data is ingested or analyzed, it is little help if the decision maker don't trust the input or the output. For instance, in the USA alone, poor data quality cost the US economy around USD 3.1 trillion a year [2].

In addition to the data collected by government agencies, some villages, particularly in Java, have developed information system containing various village data. Village Information System was developed to facilitate the service of the village officials to its citizens. It may contain population data, poverty data, number of infant and so on. In addition, it can be used for issuance of certificate, fact news and many more.

The excellence of Village Information Systems compared to government data is that these data are updated every time needed, based on the needs of the village. The data contained in Village Information System are participatory in nature. Moreover, Village Information System can also be used to deliver villages' potency, resources availability and their products. The major weakness, these data are not yet integrated with government data. Even worse, each village has their own information system which differs among villages.

2. Materials and Method

2.1. The Role of Data and Knowledge

2.1.1. Data, Information and Knowledge

Data and information are the basis for knowledge formation. Knowledge can be defined as information that valid, contextual, relevant and actionable. In simpler definition, knowledge is codified information. Moreover, knowledge can also be treated as an object that can be acquired, integrated and disseminated as commodity and becomes tradable product. Knowledge is also considered as information with meaning [4]. Furthermore, Rajpathak and Narsingpungkar [3] mentioned that distinction should be made clearly between data, information, knowledge and wisdom. Data is a set of discrete and objective facts about events. Information is message, in the form of document, picture and audio-visual. Knowledge is broader, deeper and richer than data or information since it emerges from the application, analytics and productive use of data and/or information. According to Erickson [5], data are simply observation, whilst information is data in context and knowledge is information subjected to experience, reflection on some other practice providing a deeper understanding.

The social and economic data become valuable asset when data are transformed into knowledge (gaining insight) and used for decision making for taking action. In order to make data available as knowledge, it should follow a process that take into the different phases through which data is transformed to finally lead innovation. The early phase starts with datafication and data collection. These are the activity of data generation, through the digitalization of media, and monitoring the real word (online and offline) activities generated by human and sensors. Data in this state has no inherent meaning. Data is typically useless since no information is obvious. Therefore, data analytics refers to a set of techniques and software tools that are used to extract information from data. The knowledge base refers to the knowledge that individuals or systems (organizations) accumulate through data analytics over time. The knowledge base is the precious assets and therefore need protection through

legal or technical means. The final phase is data driven decision making based on the accumulated knowledge base [1].

2.1.2. Open Data

Knowledge is the core of development in agriculture. When we discuss data in agriculture, it is not limited only on agricultural input or production, but also data for food security, nutritional levels, weather data, price, and so on. Data in agriculture is not structured such as numeric data, image data, location, and spatial data.

Data openness (open data) is crucial for agricultural development, either at the micro or macro level. For example, at the national level aggregate production and investment data can be implemented quickly if data are available. With a better access to data such as production data may help decision makers to set a target of national production. Meanwhile, at micro level, data openness may assist farmers managing his farm business. Farmers may know in advance for any threat such as drought, pests, flood that may come. Moreover, they have a better plan to sell his farm product and able to trace their product in the value chain. Farmers themselves may enter and collect data. The challenge for data openness is accuracy which needs updates regularly and level of trustworthiness for reliable data. According to GODAN [6, 7], open data are characterized by accessibility, which usually means published on the web, availability which readable by machine-readable format. Moreover, open data is characterized by open license that permits anyone to access, use and share it.

There are three specific ways open data can help solve practical problems in agriculture: (i) it will enable more efficient and effective decision making; (ii) it will foster innovation that anyone can benefit from; and (iii) it will drive organizational and sector change through transparency. Open data will lower direct cost for accessing data. Moreover, open data will drive better decision making and improved products and services based on lower cost and data availability [7]. Yet, open data also have some drawback such as privacy and security concern. Other concern is potential abuse and the danger of manipulation [2].

2.1.3. Big Data

Big Data refers to datasets whose size is beyond the ability of typical database to capture, store, manage and analyze. Big data is not only tabular, it is also includes documents, emails, video, sound bites, social media extracts, pictures, and other form of information that is difficult to fit in traditional database tables based on rows and columns [8]. Therefore, special tools and treatments are necessary to extract knowledge from it. Big data can produce minimum two types of values to an organization.

Firstly, it can be a source of innovation. Specially, it can enable development of new products, processes and services. Secondly, use of various analytics on big data can generate knowledge and insights that can support and improve organizational decision making significantly [3].

IGEL [2] said that big data is characterized by three components: (i) volume (how big the data is); (ii) velocity (how fast the data is being collected and ingested, integrated and analyzed for use in real time, virtually instantaneously); and (iii) variety (how diverse the data being collected is). Data is unstructured, which means it lacks a common format. Further, Oracle adds one dimension of the big data, ie. value [4].

2.2. SIDeKa

SIDeKa was proposed to integrate data obtained from many state institution (official data) and data already collected by communities. SIDeKa was initiated by BP2DK (Badan Prakarsa Pembangunan Desa dan Kawasan/Regional and Village Development Initiative Board) in Jakarta. Some universities are involved in this project such as Bandung Institute of Technology, Janabadara University Yogyakarta, and Atmajaya University Yogyakarta. SIDeKa was developed as the mandate of Village Laws Number 6/2014, article 86 which says that village entitled to access information and government must develop an information system in rural area which includes hardware, software, network and human resources. Since SIDeKa involving many stakeholders, any coordination should be arranged in advance, including state institution such as Ministry of National Development Planning, Ministry of Communication and Information, Ministry of Agriculture, Indonesian Statistical Bureau, and Ministry of Home Affairs.

SIDeKa is not just writing software for integrating data, but it is an information-system development related to many social work and legal aspects. The challenges are many. At the state level, many question arise, such as who will be responsible to manage data, how to minimize data redundancy which is scattered among many institution, where the server should be located, who can access the data, what kind of data to be public, how to coordinate many institutions, who will be responsible for funding, what kind of law which regulate data collection and maintenance and many more. At grass root level, there are also many village information systems which are independent and mutually exclusive each other. Moreover, there are so many actors and NGOs who already involved in the Village Information Development and they will not automatically follow SIDeKa scheme.

Based on those facts, some characteristics of SIDeKa are defined as follows: (i) SIDeKa should be based local wisdom; (ii) SIDeKa is based on participatory approach, managed by local people through deliberation. SIDeKa should not burden the village government and should meet community needs. Not all data need to be managed
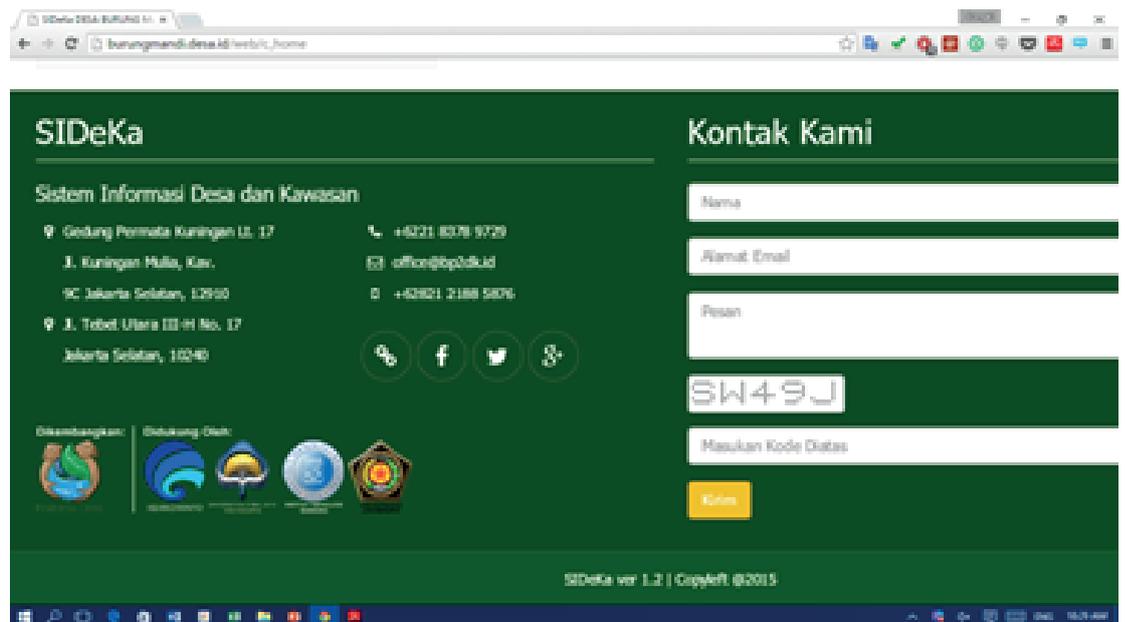
Figure 1: Homepage of SIDeKa.

by village officer; for instance, farms data are managed by farmer group; (iii) SIDeKa encourages local people to have sovereignty on their own data; (iv) SIDeKa should accommodate people interest; (v) SIDeKa should provide information to the village officer such as the excluded group in the community; and (iv) SIDeKa should provide information on village potential [9]. Kusdarjito et al [10] mentioned that SIDeKa should not replace the system already developed by village or NGOs at the grass root level. In addition, SIDeKa should be compatible with the existing data, especially for the state data. Data should be secure enough, adaptive and modular. Moreover, SIDeKa should be able to be deployed in the area with or without internet access and it should be affordable and uses open sources. SIDeKa is copyleft and some data are open to the user group and the software can be replicated freely by village.

In terms of data properties, data in SIDeKa can be categorized into three levels. First level data in SIDeKa is population and legal status of villages people, health condition, education level, number of workforce, marriage status, number of unemployment, people with disability, number of widow, number of people (family) living in poverty, type of housing and any data which is specific to each village. Second level data consist of village development plan (annually and five yearly), and village budget. The third level data consist of agricultural data, natural resources availability, economic activities, and public service provided by village. Forth level data dealing mainly with unstructured data, including messages, spatial data, pest invasion, weather data, image, news, and any data required by village which is local specific [11].

2.2.1. Implementation

SIDeKa implementation (Figure 1 and 2) should be based on regulation. The existence of SIDeKa is strong if supported by law which binding the parties involved in it. Socialization of SIDeKa is carried out to give better understanding on the benefit that will be obtained by community and village government when implementing SIDeKa. Socialization can be done by County (District) Government, facilitators or other parties who concern on SIDeKa. Better understanding on SIDeKa is very important in the process of SIDeKa implementation. Through SIDeKa village community can play an active role in the process of data collection and providing information.

Because SIDeKa involving many parties, it requires a team to execute SIDeKa. In the formation of the team, there should be a representation of the Village Government and villagers. The team formation should give special attention to the regional and sectoral representation. Regional representation ranges from hamlet to neighborhood groups, whilst sectoral representation consists of a group of farmers, fishermen, PKK and others. In addition, the team formed should pay attention to the role of each of the members, such as who will be responsible to data collection, data entry and system maintenance.

Next, the team that had already formed was trained. Some of the material provided is data collection, data entry and input mapping for spatial data, training for operators, writing course for village journalism. The training is carried out to enhance the capacity of SIDeKa operators so that the process of implementation of SIDeKa can be as good as expected. Deliberation and discussion on the data needed at the village level is intended to determine what kind of data that will be entered into SIDeKa.

Mapping is carried out because each village characteristics varies, eg. data needed for coastal villages will differ from village based on agriculture. Further, data verification is carried out to make certain that the collected data is in accordance with the actual data. Verification can be done with deliberation, such as data on poverty. After the verification is done, data can be entered into SIDeKa by operators. However, data updates must be carried out periodically, such as data of birth, mortality and other health data.

SIDeKa is still in the early stage of development. SIDeKa has been deployed in some districts, such as Kulon Progo, Pemalang, Tasikmalaya, Belitung Timur, Kota Gorontalo, Boalemo, Gianyar, dan Raja Ampat and the number of District involves in SIDeKa still counting. Yet, due to modularity and data openness in SIDeKa it will lead to data growth and accumulation which in turn demanding big data analytics.
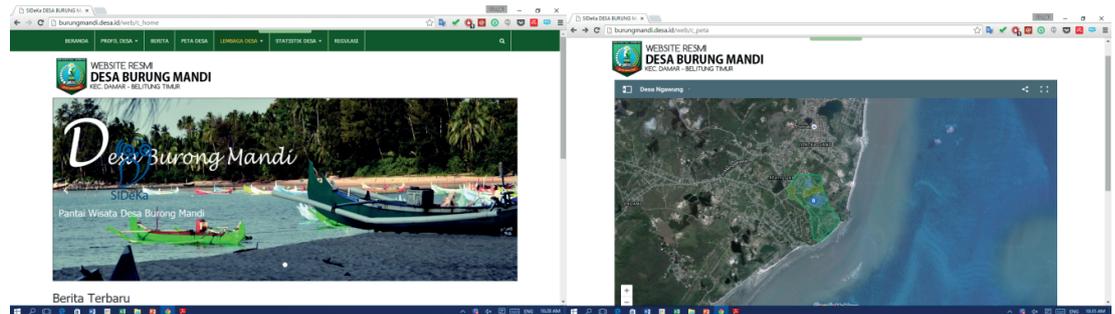
Figure 2: SIDeKa Implementation.

2.2.2. SIDeKa and Knowledge Creation: Complexity Approach

At the village level, SIDeKa will provide data, information and knowledge if used properly. For instance, Village Data can be delivered instantly by SIDeKa when it needed. Moreover, SIDeKa will provide information of poverty level or people who eligible to obtain supporting funds (or in kind). Since people who classified as poor has been agreed upon themselves, conflict that may arise can be minimized. In terms of knowledge creation, SIDeKa may provide information which can be transferred as knowledge, such as commodity price, weather data, soil data, pest invasion, so farmers can anticipate in advance.

For the supra-structure level, data will grow massively since there are more than 73,000 villages in Indonesia. Moreover, data fed to SIDeKa system are unstructured, for example image or voice. Therefore, big data analytic should be used to interpret and harvesting data in SIDeKa at the supra-structure level. Since SIDeKa involves many stakeholders and bottom up in nature, systems approach in development should be used such as D3 (Doing Development Differently). Classical approach will not sufficient to handle such huge complex system.

In the futures, two questions may arise. First, who will responsible for the main and secondary server. Who will become "super admin" and who will ensure users access to the server. Government think tank bodies such as Presidential Advisory Board and Universities should be granted for access, especially for macro data whilst NGOs, farmers, universities on micro data. Universities and NGOs may provide additional data based on their research.

3. Conclusion

The success or failure in implementing SIDeKa depends on the government goodwill. SIDeKa should be treated as complement for official data obtained by Indonesian Statistical Bureau, yet the velocity and variability (type of data) in SIDeKa is greater than those in official data. SIDeKa can be used instantaneously by its user. Moreover,

indicators to measure the success or failures of SIDeKa should be developed, but it must be based on system approach which are more adaptive.

References

[1] OECD: Interim Synthesis Report Data-Driven Innovation for Growth and Well-being, Oct 2014.

[2] IGEL (Initiative for Global Environmental and Leadership: Special Report on Sustainability in the Age of Big Data), Wharton University of Pensylvania, Sep. 2014.

[3] Rajpathak, T. and A. Narsingpungkar. 2013. White Paper: Managing Knowledge from Big Data Analytics in Product Development. Tata Consultancy Services.

[4] Carayannis, E. and N. Kabir. 2013. Big Data, Tacit Knowledge and Organizational Competitiveness. Journal of Intelligence Studies in Business 3, 54-62.

[5] Erickson, S. and H. Rothberg. 2014. Big Data and Knowledge Management: Establishing a Conceptual Foundation. The Electronic Journal of Knowledge Management 12 Issue 2 (pp108-116) available online at www.ejkm.com

[6] GODAN (Global Open Data for Agriculture and Nutrition). 2015. How can We Improve Agriculture, Food and Nutrition with Open Data, Open Data Institute. 2013.

[7] GODAN (Global Open Data for Agriculture and Nutrition). 2015. How can We Improve Agriculture, Food and Nutrition with Open Data, Open Data Institute. 2015.

[8] Albala, M. 2011. Making Sense of Big Data in the Petabyte Age. Cognizant, New Jersey.

[9] Suharyana. 2015. General Guidance for Organizing of SIDeKa: Village and Regional Information System. In Indonesian. Prakarsa Desa, Jakarta.

[10] Kusdarjito, C., I. Wisnubhadra, B. Suhardiman, Suharyana, I. Meilandi, Purwoko, Sobirin and A. Triadi. 2015. General Guidance for Organizing of SIDeKa: Village and Regional Information System. In Indonesian.