

KNOWLEDGE GRID MODEL IN FACILITATING KNOWLEDGE SHARING AMONG BIG DATA COMMUNITY: SYSTEMATIC LITERATURE REVIEW

¹SARA HOSSEINIOUN, ²RUSLI ABDULLAH, ²YUSMADI YAH JUSOH and ²MARZANAH JABAR

Faculty of Computer Science and Information Technology, University Putra Malaysia, Selangor, Malaysia

¹sarahosseinioun@gmail.com, ²{rusli, yusmadi, marzanah}@upm.edu.my

ABSTRACT

This systematic literature review (SLR) aimed to identify issues of knowledge sharing in big data community as a huge amount of data and highlight the knowledge grid's characteristics which can improve knowledge flow between members of a community. In achieving the objectives, this study follows Kitchenham's, (2009) method for SLR. In consequence, three research questions have been chosen in the first phase and in following phases all the related article which help to answer the research questions by the keywords that have been collected. Analysis of results revealed the difficulty of knowledge sharing between community members' interaction and collaboration and there is a gap in researches on big data user's roles on sharing through the big data community. Moreover, knowledge grid as communication infrastructure can effectively acquire, represents and exchanges massive amount of knowledge however current knowledge grid models focus on knowledge discovery through big data and there is a lack of efficient knowledge grid model for distributing knowledge among community members by considering their influence on knowledge.

Keywords: *Big Data, Big Data Community, Knowledge Management, Knowledge Sharing, Knowledge Grid.*

1. INTRODUCTION

The term of Big Data is being increasingly used online, offline and almost everywhere on the planet. It comes under the blanket of Information Technology, which is now part of about all other technologies and fields of studies and businesses. In short, while technology revolution makes generating enormous data simple for digital devices users all over the world and this phenomenon has extended swiftly and increasing scale which called big data. The scale is not the only issue for managing these massive generated data since it involves a great variety of data forms such as text, sound, video etc., and time as another dimension of data stream [1]. Also, traditional ways can't support the huge amount of knowledge flow between big data users such as individuals and organization as members of big data community at real time while sharing knowledge is the significant feature of big data community.

The Knowledge Grid as intelligent and sustainable environment enables users to publish, share and manage knowledge effectively. [2] The

Knowledge Grid will be a mechanism that can synthesize knowledge from data through mining and reference methods and enable search engines to make references, answer questions, and draw conclusions from masses of data. [3] Therefore Knowledge Grid will develop research toward the next-generation Web, using it to build a more efficient and effective intelligent application platform.

This SLR provides included literature on sharing knowledge in big data community and knowledge grid to answer the research question and review the previews studies to find out the relationship between research keywords from the bases. Three research question as RQ1: What is big data community effects knowledge management in concern of knowledge sharing? RQ2: How does knowledge grid support knowledge management specifically in knowledge sharing? RQ3: How does knowledge grid apply in big data community? have been designing at first phase of research method and SLR follow the path to answer them. At the last session, the result of this research has been categorized by research questions.

2. BACKGROUND AND RELATED WORKS

This studies' intention is to find out the current situation of knowledge sharing in big data and knowledge grid and analysis their relationship. Thus, making a clear understanding of big data, big data community, knowledge management, knowledge sharing and knowledge grid are necessary to analyze their relationships more accurately.

2.1. What is Big Data?

Big Data knew as large volume of data which can be structured or unstructured and generate by devices, systems or users. These structured data are related to schemas are known, and resident in identifiable repositories, such as databases; and "Unstructured data" may include web pages, blogs, news and social media, repositories of texts such as publications, internal organizational knowledge bases, emails, videos, photos and a host of user-generated content [4]. Big Data has been characterized in terms of its volume, variety, velocity, variability, value, and veracity.

- i. *Volume*: The word "Big" in Big data itself defines the volume as a huge amount of data are generating every second [5], e.g. Facebook daily generates over 500 terabytes of data, and Wal-Mart collects more than 2.5 petabytes of data every hour from its customer transactions [6]. Excessive data volume creates issues including storage, how to determine relevance in the large volumes of data and how to create value from the data that is relevant [7].
- ii. *Variety*: Data being produced is not of a single class as it not only includes the conventional data but, also the semi-structured data from different resources like web Pages, Web Log Files, social media sites, e-mail, documents, sensor devices data both from active and passive devices. All this data is total dissimilar consisting of raw, structured, semi-structured and even unstructured data which is hard to be handled by the obtainable traditional analytic systems. [8]
- iii. *Velocity*: It Deals with speed data comes from different sources and moves around [5], For example, the data from the sensor devices would be continuously moving to the database store and this quantity won't be small enough. Thus, systems are not capable enough of performing analytically on the data which is continuously in motion [8]. The rate of change in the data and how quickly it must be used to create real value is what velocity refers to.

Traditional techniques are poorly suited to storing and using high-velocity data [7].

- iv. *Variability*: It considers the inconsistencies of the data flow. Data loads become tough to be maintained particularly with the rise in the usage of the social media which normally causes a peak in data loads with assured events happening. [8] It is an extremely essential feature which often confused with variety and the best example of it is Google or Facebook repository stores and generates many different types of data. From this variety of data type, one of them is brought to use for mining and making sense out of it by users but every time the data offers a different meaning whose meaning is constantly and rapidly changing [6].
- v. *Value*: It's power and leverage of the data [5], For example, significant values can be extracted from the stream of clicks left behind by the internet users and this is becoming a backbone of the internet economy [6]. The user can be able to run certain queries against the data saved and thus, can abstract essential results from the filtered data obtained and can also order it according to the magnitude they need. This information helps them to establish the business trends according to which they can alter their strategies. [8]
- vi. *Veracity*: The trust worthies, quality or accuracy of the date mentioned as veracity characteristic of big data [5]. This feature suggested by IBM and it is more about understanding the data, as there are integral discrepancies in almost all the data collected. It represents the untrustworthiness inherent in many sources of structured as well as unstructured data [6].

2.2. What is Big Data Community?

The community is a small or large social unit, who have something in common while the virtual community-based approach is an effective way of sharing knowledge and facilitates informal sharing of the available knowledge from experienced and skilled users with other members of the community [9]. Moreover, accumulating social capital, encouraging collective trust and receiving social support facilitated for community members in online communities as a social group, while users or consumers interact with each other on the internet [10]. In other words, the internet supports collaborative groups where marketers and consumers interact to develop more engaging products and services. These unions are usually

based on particular common interest and their members are linked by a shared respect to that mutual interest. Thus, by expanding the above explanations, big data community have been defined as a strong relationship between community members as a decision maker, service providers or social media users which can be individuals or organizations while huge amount and variety of storage knowledge flow among them around the world. These members join the community from various areas such as health and human welfare, business and commerce, government and public services, science and education, social media and etc. to discuss, give advice, collaborate and exchange knowledge and sources to improve their decision making, services, and product.

2.3. Knowledge Management and Knowledge Sharing

Knowledge management focuses on managing knowledge as actionable information in terms of knowledge creation or acquisition, storage, distribution and application [11]. Fundamentally it is about making correct knowledge from available and reliable knowledge source to authorized user at the right time.

In general, users share knowledge while they want to solve problems, discuss certain topics and give their comments on others’ opinions. Darr and Kurtzberg (2000) claimed that the knowledge sharing is a process which means to gain experience from others. The knowledge sharing process involves knowledge transferring and knowledge receiving [12]. Knowledge transferring is the dissemination of personal ideas, techniques or know-how. Knowledge receiving means to acquire knowledge [13]. Therefore, knowledge sharing can be defined as the sharing of community-related information, ideas, suggestions, and expertise among individuals [14].

To analyze the knowledge sharing current situation is must understand the factors affects knowledge sharing and motivate users to exchange their knowledge with each others. Table 1 illustrates factors which influence knowledge sharing and categorizes them to Behavioural, Environmental and Technical factors as a term of users that generate big data, organization which make structure rules and ICT technology.

Table 1: Three Category influence knowledge sharing.

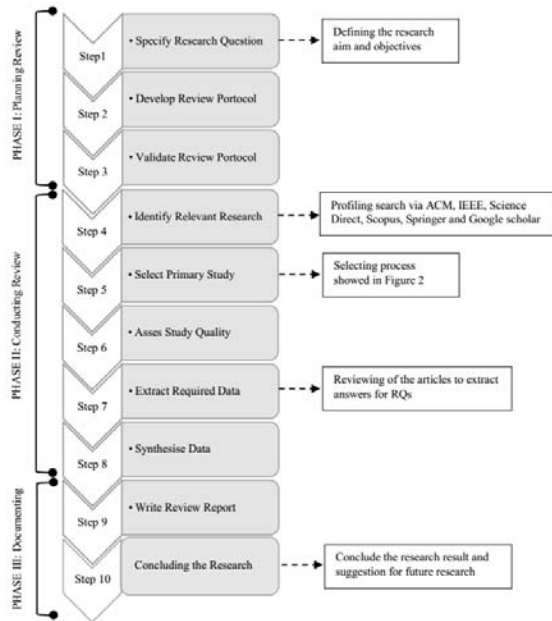
Categories		Factors
Behavioral	The characteristic that leading to a pleasure in helping people without asking for a return and mutually contingent exchange of benefits and a set of sentiments associated with mutual gratification [15].The degree of belief in good intentions, benevolence, competence, and reliability of members sharing knowledge [16]. Treat the other colleagues as his/her own fellows, and hold faith in each other [15].	Reliability, Validity,
Environmental	The factors mainly include formalization, complexity, and centralization. Formalization refers to the extent to which work activities are bound by the company’s formal rules, regulations and procedures. Complexity refers to the extent to which duties are segregated in a job. Concentration refers to the distribution of the decision-making power on work activities [17]. Also, it is a construct that reflects the dynamic, personal energy with which an action is performed. When focusing on knowledge sharing, motivation can be defined the inner drive to share knowledge with a coworker [18].	Reliability, Security, Privacy
Technical	The extent to which team members have confidence and trust in each other’s knowledge. The extent to which communication among developers is timely and relevant to development tasks [19].	Accessibility, Security

2.4. What is Knowledge Grid (KG)?

The knowledge grid is a communication infrastructure which obtains, represents and exchanges effectively massive data and information while integrates and converts them into useful knowledge through mining and reference methods [20]. Knowledge Grid is designed on top of computational Grid mechanisms provided by Grid environments [21]. knowledge grid offers high-level tools and techniques for the distributed, mining and extraction knowledge from data repositories available on the Grid [22]. In addition, knowledge Grid is an intelligent, sustainable internet application environment that enables users to effectively capture, publish, share, and manage explicit knowledge resources [2]. Thus, Knowledge Grid uses Grid Computing for its high storage capability and processing power [8]. Grid computing is a model of distributed computing that uses geographically and administratively disparate resources and individual users can access computers and data transparently, without having to consider location, operating system, account administration and other details while the details are abstracted and the resources are virtualized [7]. Knowledge grid had been designed on computational grid mechanisms as infrastructure that provides dependable, consistent, pervasive and inexpensive access to high-end computational capabilities. It provides secure access to high ability of shared processing power suitable for high throughput applications [7]. Moreover, it is an evolved form of data grid which is an infrastructure to support data storage, data discovery, data handling, data publication, and data manipulation of large volumes of data stored in various mixed databases and file systems [7].

3. METHODOLOGY

This SLR follow Kitchenham, (2009) method to conduct Systematic Literature Review, which represents three phases as Plane Review, Conduct Review, and Document Review as systematic literature review process that includes ten steps [23]. Figure 1 represents the systematic literature review process.



3.1. Plan Review

This SLR research aims to evaluate the existing research published on Knowledge Grid and Big Data by employing an established profiling approach to investigate and analyze different Knowledge Management, Knowledge Sharing in big data community and Knowledge Grid methods and or approaches. Table 2 illustrates the research questions have been designed for this study and their motivations.

Table 2: Research Questions

Research Question	Description and Motivation
RQ1: What are big data community effects knowledge management in concern of knowledge sharing?	This question aims to identify the current situation of knowledge management and specifically knowledge sharing in big data community and distinguish the ineffectiveness and issues.
RQ2: How knowledge grid support knowledge management specifically knowledge sharing?	The purpose of this question is to find and analyze the capability of knowledge detect and share in grid platform.
RQ3: How does knowledge grid apply in bid data community?	This question intends to analyze how knowledge grid can apply to big data community while contemplating big data characteristics and review current knowledge grid models in big data community.

This systematic search started with a developed comprehensive review protocol based on [23] guiding principles and procedures of the

systematic literature review. This protocol identifies the review background, search strategy, research questions, data extraction, criteria for study selection and data synthesis. The research questions and the background of this review are described above, although more details about other elements have been provided. The review protocol does not only help to increase the accuracy of the review also to reduce researcher bias.

The purpose of setting up the criteria for inclusion and exclusion is to make sure, only relevant researches to this study have been used. [24] consider inclusion and exclusion criteria which have been pursued in this SLR. This study considers research articles (from journals, conferences, and workshops) in the English language, published from January 2000 to 2017 in digital databases and eliminates editorials, prefaces, poster sessions, panels and tutorial summaries. Also, duplicated studies and articles which were not related to the research question or their full text not available have been excluded. Table 3, shows a summary of these criteria. Further, when different versions of an article exist, which may appear as a book chapter, conference or journal article, only the complete version of the article is included and the others are excluded.

Table 3: Inclusion and Exclusion Criteria

Included articles are:	Excluded articles are:
Full text	Full-text is not available
Published from 2000 till 2017	Non-English papers
Related to the research question	Duplicated papers

Initially, 373 journal articles were identified from the ACM, IEEE, Science Direct, Scopus, Springer database and relating to articles published during the period from 2000 to 2017. After assessing the 337 articles (from refereed journals), 212 papers were discarded, and finally, 125 papers were selected and taken forward for further interrogation which from them 46 paper left after quality assessment. Figure 2 illustrates the articles selection process.

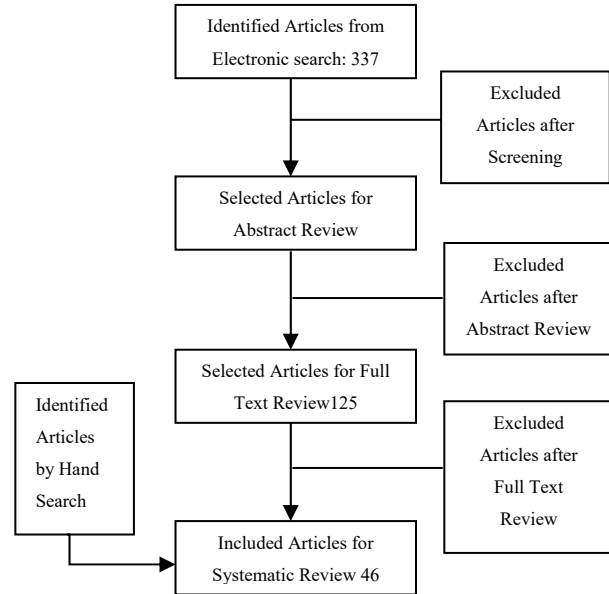


Figure 2: Article Selection Process

The final selected articles which count as 46 for this SLR all full text reviewed by considering their connection to research questions and listed in references.

3.2. Conduct Review

To identify the relevant articles through the selected online databases, the following keywords have been searched.

- Knowledge Grid, Knowledge Grid Environment, knowledge Grid Model, Knowledge Sharing in Knowledge Grid, Big Data, Big Data Challenges, Knowledge Sharing in Big Data OR Problems OR Barriers OR Methods, Knowledge Management in Big Data, Big Data Community.

Five online libraries, ACM, IEEE, Science Direct, Scopus and Springer, including google scholar have been reviewed and articles and conferences journal contain keywords in their title, abstract and keywords have been collected for future review.

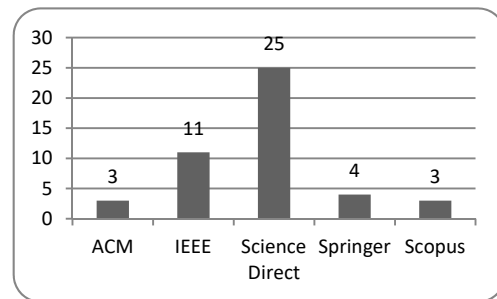


Figure 3: Primary Studies by Online Library

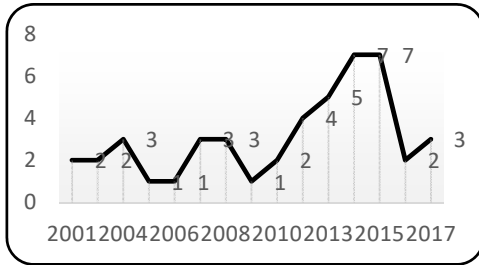


Figure 4: Primary Studies by Year

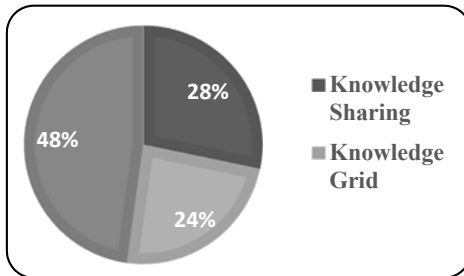


Figure 5: Primary Studies by Main Keywords

In addition, as Figure 3 shows most of the collected articles by relevancy which follow quality criteria belong to science direct and IEEE online libraries while they have been presented mostly after 2013 to recent time that illustrated in Figure 4. Also, most of these researches focused on big data around 48% compared to other keywords, knowledge sharing and knowledge grid which represent in Figure 5.

The selected papers were analyzed and synthesized before using for findings and results by quality assessment screening. The quality assessment aimed to evaluate the completeness resources and avoid any bias to maximize the validity of SLR. Thus, 4 questions (Q1-Q4) have been chosen to assess completeness, relevancy, and credibility of selected articles which have been presented in Table 4. Each question has only 3 answer options: Yes=1; Partially=0.5; and No=0. All the selected papers have been evaluated by the by these quality criteria and it result categorize by the studies related to each research question and illustrated in Tables 5, 6 and 7 which shows the redundancy of paper and interconnection of research questions.

Table 4: Study Quality Assessment Criteria

NO	Quality Assessment Questions	Answers
Q1	Is there a clear description of the aims and objectives of the investigation?	Yes/No
Q2	Is the paper explained the method of analysis pertinent and adequately?	Yes/No/Partially
Q3	Is the paper supported by primary data?	Yes/No/
Q4	Is the paper explained the model structure in detail?	Yes/No/Partially

Table 5: Quality Assessment Result for Research Question 1

References	Q1	Q2	Q3	Q4	Total
[6]	1	1	1	0	3
[25]	1	1	1	1	4
[7]	1	1	1	1	4
[26]	1	1	1	1	4
[1]	1	1	1	0	3
[27]	1	1	1	1	4
[28]	1	1	0	1	3
[10]	1	1	0	1	3
[29]	1	1	1	1	4
[30]	1	1	0	1	3
[19]	1	1	0	1	3
[8]	1	1	0	1	3
[2]	1	1	1	1	4

Table 6: Quality Assessment Result for Research Question 2

References	Q1	Q2	Q3	Q4	Total
[8]	1	1	0	1	3
[22]	1	1	1	1	4
[7]	1	1	1	1	4
[31]	1	1	1	1	4
[32]	1	1	0	1	3
[33]	1	1	1	1	4

Table 7: Quality Assessment Result for Research Question 3

References	Q1	Q2	Q3	Q4	Total
[8]	1	1	0	1	3
[34]	1	1	1	1	4
[33]	1	1	1	1	4
[11]	1	1	1	1	4
[7]	1	1	1	1	4
[35]	1	1	0	1	3
[36]	1	1	0	1	3

3.3. Document Review

The research has been defined the keywords and their relationship and build the fundamental of research step by step through answer research questions. The finding of this study has been representing as SLR result and analysis session and have been categorized by research questions.

4. RESULTS AND ANALYSIS

The main aim of the research was to answer and analyze the research questions in reviewed articles which categorize at this session.

4.1. What is Big Data Communities Effects on Knowledge Management in Concern of Knowledge Sharing? (RQ1)

Since data characteristics of Big Data community as a small or large social unit, accessing, managing and governing data and share it between members will be challenging; While data warehouses store huge amounts of sensitive data such as financial transactions, products and services details, medical procedures, insurance claims, research results, diagnosis codes, personal data, etc. Organizations and businesses need to ensure their privacy and their security infrastructure that enables employees and staff of each division to only view and access relevant data for their department [6]. Sharing data and information needs to be balanced and controlled to maximize its effect, while organizations store large-scale datasets which poses an enormous task of sharing and integrating key information across them and establishing close connections and harmonization with their business partners [25].

The issue reveals when the amount of amassed data that is becoming so large which

finding the most valuable pieces of information complicated. Organizations have been limited to using subsets of their data and constrained to simplistic analyses because the sheer volumes of data overwhelm their processing platforms [7]. The huge size of data needs more workload security for sharing. Otherwise, most Big Data are stored in a distributed way, and the threats from networks also can aggravate the problems [26]. Rapid processing of a large number of metadata records and datasets to satisfy large number users' will be difficult. Also manage the increasing rate of data flows for highly heterogeneous data models, encoding formats, and access service interfaces are challenging [29]. Enormous data from different sources, type, and process greatly interconnected, interrelated is complicated to analyze, manage and share at the real-time [1]. There are security issues when sensitive data are transmitted from a data owner's local server to a big data platform; and, there are sensitive data computing and storage security problems and issues involving secure data destruction [27]. The reliability of data is a challenge to deal with, to be able to make justified decisions based on the data. The less representative the data is the less valuable is it as well. Lai and Hsiao (2014) stress the importance of the validity of data as data should have enough quality and be appropriate for the purpose such as decision-making by users and organizations. Because of the amount of data and their complexity, it is important to be aware of the data and clarify that they are comparable when different datasets are merged together [28]. Table 8 summarize managing knowledge key challenges through big data community.

Table 8: Knowledge Sharing Challenges among Big Data Community

	Accessibility	Reliability	validity	security	privacy
[6]		X	X	X	X
[25]	X	X			
[7]	X		X		
[26]				X	X
[1]	X	X			
[27]		X	X	X	
[28]		X	X		X
[10]	X	X		X	
Total	4	6	4	4	3

Big data collected by various devices around the world and arranged in different areas need a highly decentralized cyber-infrastructure for processing and Sharing. Must attention, cyber structure and rapid development of internet involve

worldwide and communication techniques to support different type of applications and use globally distributed digital resources. Also, they provide data acquisition, storage, management, integration, processing, and utilization for researchers to conduct trustworthy exploration.

In combine of community member's such as individual users or organizations and technical role on knowledge sharing among big data community and current situation of knowledge sharing; facilitating knowledge sharing involve improving accessibility, reliability, validity, security, and privacy of shared knowledge in the community. **As mentioned before factors influence knowledge sharing categorized from three aspects of user's behaviors, environment and technical which each has direct impacts in facilitating knowledge sharing factors while the relationship is an open field for research in big data community.**

4.2. How Knowledge Grid Supports Knowledge Management Specifically in Knowledge Sharing? (RQ2)

Zhuge (2004) highlighted some characteristics for knowledge grid which show its efficiency in knowledge sharing. First, it makes users able to access and manage the distributed knowledge from all over the world without knowing the knowledge location through a single entry. Also, distributed knowledge can be clustered by their relevancy, through using Metaknowledge, to provide reasonable and explainable knowledge service. Finally, knowledge would not statically store in knowledge grid environment and it can dynamically evolve to keep up-to-date. So, knowledge grid can improve itself during use. Since the last decade there has been a substantial increase in computing and network performance, mainly because of faster hardware and sophisticated software, thus, these commodities technologies and fast networks have been used to develop high-performance computing systems, called clusters, to solve resource intensive problems in several application domains. But these systems have been found incapable of handling massive data processing and storage. Therefore, for such challenges which revolve around data managing its access, distribution, processing and its storage, computational infrastructure, coupling wide-area distributed resources such as databases, storage servers, high-speed networks, supercomputers and clusters for solving large-scale problems have been developed known as Grid Computing. Another reason for the growth of grid computing is the potential for an organization to reduce the capital

and operating cost of its computing resources while maintaining the computing capabilities it requires. This is because the computing resources of most organizations are vastly underutilized, but are necessary for certain operations [7]. The Grid is efforts to create an advanced cyberinfrastructure, aiming at an adaptive wide-area resource environment, integrating higher-level services that enable applications to adapt to heterogeneous and dynamically changing meta-computing environment, with ease, low cost, reliability and regardless of the location and device. [32] Whereas knowledge Grid provides a distributed system which connects sources dynamically and its essential analyze factors that affect knowledge sharing while these factors directly related to individual users and organization as members of big data community.

Requirement, resources, roles, and rules are the principle of Knowledge Grid with machine-understandable semantics, a resource can actively and dynamically gather relevant resources and fuse them to provide appropriate on-demand services for applications by understanding requirements and functions and relating them to each other. So, to have accurate knowledge and avoid tampering and reply in storage knowledge valid and trustful knowledge and an accurate semantic relationship between them is essential. [31] Trustworthy and validity of received knowledge modify knowledge sharing among the community. That can cover organizational rules that affect knowledge sharing and cover the environmental factors which influence knowledge sharing. In addition, Knowledge Grid focused on three benefits which are, the public can share with other grids; second, the privet group knowledge can only share by the same group and final the privet knowledge can only use by its owner [33]. Data flowing across the different nodes of the grid is so much value for its owner, so it should go only to those who are intended to receive it [7]. Also, increase user satisfaction and motivate community members to flow share knowledge with other members.

4.3. How Does Knowledge Grid Apply in Big Data Community? (RQ3)

Grids were originally designed for dealing with problems involving large amounts of data or compute-intensive applications [34]. Knowledge Grid as a grid-based architecture which supporting distributed knowledge must be able to link different resources into the grid and use resources to perform some tasks while composing grid sources to form newly combined resources [33]. Thus, it should

perform knowledge discovery process and distributed data analysis tasks in high variety, velocity and higher volume of data in big data community. The challenge is to filter the most significant data from all the data collected by different users and stored in different servers in big data chaos by community members request and contribute it. Also, data as a set of direct facts after process get the meaning and become information. Information justifies and develops into relevant and actionable knowledge to solve and answer the question. [11]In Grid number of servers are interconnected by a high-speed network, which each server plays one many roles through the grid. It will help to increase storage capability which is so important in big data [35]. The main duty is the find valuable data, information and most significant knowledge of raw storage data, which help flowing knowledge among members of big data community more effective. Table 9 represents the current designed knowledge grid model and the factors they focus and the methods they use to achieve their result.

Table 9: Current Knowledge Grid Models in Big data

Author	Method	Focus
[7]	Distributed catching as a technique to store data in memory	Accessibility, Reliability, and Security
[36]	Ant-based algorithm	Knowledge Clustering, Accessibility, and Reliability
[33]	Three-dimensional knowledge space	Categorize Knowledge, Accessibility, Reliability, and Validity

The methods preview models adopted, combined with grid computing and most of their attention is on knowledge discovery and next knowledge sharing, so they couldn't cover all the knowledge sharing facilitating factors. The focus of current models as mention is on detecting knowledge in big data while in community members impact on knowledge activity more than other environments. In addition, they concentrated on the technical view and there is a gap of analysis about users and environment role in their suggested models. The significant feature of big data community is to distribute requested knowledge among the community by member's commands and their satisfaction which there is lack of knowledge grid model in this area.

5. LIMITATION OF STUDY

The result of analysis the related research showed there are few empirical studies directly focused knowledge grid abilities for knowledge sharing among big data community. For instance, around 30% of researches didn't support their findings with primary or secondary data and this percentage increased to 42 when comes to current knowledge grid models. As result of RQ1 present, this research is not the first to highlight the difficulty and complexity of knowledge sharing among big data, although knowledge grid as distributed infrastructure didn't consider in this need.

Therefore, beyond the analysis of what is already available about knowledge grid, should also attended to what more need to discover and identify from different viewpoints.

6. DISCUSSION

This study has been reviewed the difficulty of converting raw data to understandable knowledge and share it in big data community members. Big data community is based on regular interaction, a common objective, relationships and communication between the big data users as individuals and organizations from different areas such as Medical, Business Education and etc., thus, facilitating knowledge flow between members is significant. Big data characteristics make finding valuable information among increasing rate of data, flow and distribute it between users at real-time complicated [1]. However, as finding from RQ1 from 13 research (28.3%) revealed, there is difficulty in knowledge sharing through the big data community while its characteristics have a direct impact on knowledge accessibility, reliability, validity, security, and privacy which lack of them cause inefficient knowledge sharing among this community. So current situation of knowledge sharing in big data chaos has a significant distance from facilitating.

In addition, as reviewed in RQ2 knowledge grid as dynamic distribute system, connected knowledge around worldwide and guaranteed proper knowledge clustering as a minimum complete knowledge set for solving problems. To achieve this goal, it needs to create new knowledge organization models by finding the semantic relationship in collected knowledge which isn't statically stored [32]; 6 studies (13%) mentioned technical features impact knowledge sharing in knowledge grid by influence knowledge clustering,

relevancy, connection, availability, security, and accuracy in grid while users and organization as member of big data community have important role on knowledge activity in big data community.

However, as mention in RQ3, 7 research (15.2%) argued that current knowledge grid model for knowledge sharing more focused on knowledge clustering or categorizing in big data and its accessibility and reliability while privacy, security, and validity ignored. Moreover, there is lack of research of distributing knowledge between big data community by considering members' role in knowledge activities. Although, there is lack of comprehensive Knowledge Grid model for sharing knowledge in big data community.

This research finding showed users behavioral, organizational or environmental and technical aspect influence knowledge sharing and have a considerable effect on knowledge generate, define, aggregate, combine, connect and evaluate in big data. However, there is lack of research and analysis of these aspects through the big data community while they have a significant influence on knowledge accessibility, reliability, validity, security, and privacy which lack of them cause inefficient knowledge sharing among this community. Moreover, these aspects didn't consider in current knowledge grid models and there is a gap in them roles on knowledge grid components and distributing system. However, more focus on using knowledge grid in big data era is for knowledge discovery and there isn't a clear analysis and method for sharing massive discovered knowledge between the community member and cover their needs.

7. CONCLUSION

This systematic literature review(SLR) highlights issues of knowledge sharing in big data community as the massive volume of both structured and unstructured data which are interconnected by considering how users and organizations as the member of community effect in knowledge flow. Knowledge Grid as a dynamic distributing system which can handle massive data processing and storage facilitate knowledge sharing. So, this SLR analyses the way knowledge grid can improve knowledge sharing and current knowledge grid model in big data. It follows Kitchenham et. al, (2009) guidelines to conduct this research and design three research question to lead the researcher through the study. It finds out there is an issue in accessibility, reliability, validity,

security, and privacy of shared knowledge among the community which decreases the efficiency of knowledge sharing and user satisfaction. However current knowledge grid models focused on knowledge clustering and discovery among big data there is a lack of empirical study and a comprehensive model to consider community members and environment roles on knowledge sharing.

8. REFERENCES

- [1] Che D., Safran M. and Peng Z. , "From Big Data to Big Data Mining: Challenges Issues, and Opportunities," *DASFAA*, vol 7827, no. 1, pp. 1-15, 2013.
- [2] Zhuge, H., "China's E-Science Knowledge Grid Environment," *E - Science*, vol. 4, no. 1, pp. 13-17, 2004.
- [3] Cannataro M. and Talia D., "Semantic and Knowledge Grids:Building the Next-Generation Grid," *IEEE Computer Society*, pp. 56-63, 2004.
- [4] Das M. , Cui R. , Campbell D.R. Agrawal G. and Ramnath R. , "Toward: Methods for Systematic Research On Big Data," in *IEEE Conference Publications* 2015.
- [5] B. Marr, "5 Vs of big data," 19 March 2015. [Online]. Available <http://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters>.
- [6] Sivarajah U. , Kamal M. M., Irani Z. Weerakkody V. , "Critical analysis of Big Data challenges and analytical methods,' *Journal of Business Research*, vol. 70, p 263–286, 2017.
- [7] Chandhini C. and Megana L.P, "Grid Computing-A Next Level Challenge with Big Data," *International Journal of Scientific & Engineering Research*, vol. 4 no. 3, pp. 1-5, 2013.
- [8] L. M. Aminu, "Implementing Big Data Management on Grid Computing Environment," *International Journal of Engineering And Computer Science*, vol 3, no. 9, pp. 8455-8459, 2014.
- [9] Lin H.F. , "Determinants of successful virtual communities: Contributions from system," *Information & Management*, vol 45, p. 522–527, 2008.

- [10] Hsiao C.C. and Chiou J.S., "The impact of online community position on online game continuance intention: Do game knowledge and community size matter?," *Information & Management*, vol. 49, p 292–300, 2012.
- [11] Abdullah R.H., Knowledge management system in a collaborative environment kuala Lumpur: University Putra Malaysia 2008.
- [12] Hendriks P., "Why share knowledge? The influence of ICT on the motivation for knowledge sharing," *Knowledge and Process Management*, vol. 6, no. 2, pp 91-100, 1999.
- [13] Kim S.J. ,Hong J.Y. , Suh E.H. , "A diagnosis framework for identifying the current knowledge sharing activity status," *Expert Systems with Applications* vol. 39, p. 13093–13107., 2012.
- [14] Yu T.K. ,Lu L.C. , Liu T.F. , "Exploring factors that influence knowledge sharing behavior via weblogs," *Computers in Human Behavior*, vol. 26, p. 32–41, 2010.
- [15] Lin T.C., Wu S. and Lu C.T., "Exploring the affect factors of knowledge sharing behavior: The relations model theory perspective.," *Expert Systems with Applications*, vol. 39, p. 751–764, 2012.
- [16] Chen C.J. and Hung S. W., "To give or to receive? Factors influencing members knowledge sharing and community promotion in professional virtual communities.," *Information & Management*, vol. 47, p. 226–236., 2012.
- [17] W.-B. Lin, " The exploration factors of affecting knowledge sharing – The case of Taiwan's high-tech industry," *Expert Systems with Applications*, vol. 35, p 661–676, 2008.
- [18] Siemsen E. , Roth A.V. and Balasubramanian S., "How motivation, opportunity, and ability drive knowledge sharing: The constraining-factor model.," *Journal of Operations Management*, vol 26, p. 426–445, 2007.
- [19] Chena X. , Li X., Clark J. G., Dietrich G B., "Knowledge sharing in open source software project teams: A transactive memory system perspective," *International Journal of Information Management*, vol. 33, p. 553– 563, 2013.
- [20] Wang G. , Wen T. , Guo Q. and Ma X. "A Knowledge Grid Architecture Based on Mobile Agent," Washington, DC USA, 2006.
- [21] Nigro H.O. , Elizabeth S. and Císaro G. "Ontologies Application to Knowledge Discovery Process in Databases, Buenos Aires: IGI Global, 2009.
- [22] Cannataro M. ,Talia D. and Trunfio P. "Design of Distributed Data Mining Applications on the KNOWLEDGE GRID," *Next Generation Data Mining* pp. 191-195, 2002.
- [23] Kitchenham B. , Brereton O.P., Budger D., Turner M. , Baily J. and Linkman S. "Systematic Literature Review in Software Engineering," *Information and software technology*, vol. 51, pp. 7-15 2009.
- [24] Balaida A., Rozan M.Z.A., Hikmi S.N. Memon J., "Knowledge maps: A systematic literature review and directions for future research," *International Journal of Information Management*, vol. 36, p 451–475, 2016.
- [25] Al-Nuaimi E., Al-Neyadi H. , Mohameed N. and Al-Jaroodi J., "Applications of big data," *Journal of Internet Services and Applications*, vol. 6, no. 1, pp. 1-15 2015.
- [26] Chen C.L.P. and Zhang C.Y., "Data-intensive applications, challenges, techniques and technologies: A survey of Big Data," *Information Sciences*, vol. 275 no. 1, p. 314–347, 2014.
- [27] Dong X., Li R. , He H., Zhou W., Xue Z and Wu H. , "Secure Sensitive Data Sharing on a Big Data Platform," *Tsinghua Science and Technology*, vol 20, no. 1, pp. 72-80, 2015.
- [28] C. Fredriksson, "Knowledge Management with Big Data Creating New Possibilities for Organizations," Gothenburg, 2015.
- [29] Nativi S. , Mazzetti P. , Santoro M. Papeschi F. , Craglia M. and Ochiai O. "Big Data challenges in building the Global Earth Observation System of Systems," *Environmental Modelling & Software*, vol. 68, no. 1, pp. 1-26, 2015.

- [30] Lai M. K. and Hsiao S., "Developing data collection and management systems for decision-making: What professional development is required?," *Studies in Educational Evaluation*, vol. 42, pp. 63-70, 2014.
- [31] Zhuge H., *The Knowledge Grid* Singapore: world scientific publishing 2004.
- [32] Zhuge H. and Sun X. , "Semantics Knowledge and Grids on Big Data," *Future Generation Computer Systems* vol. 64, p. 163–164, 2016.
- [33] Zhuge H., "A Knowledge grid model and platform for global knowledge sharing," *Expert systems with Applications*, vol. 22 pp. 313-320, 2002.
- [34] Cesario E., Congiusta A. , Talia D. and Trunfio P., "Designing data analysis services in the Knowledge Grid," *CoreGRID*, pp. 1-15, 2007.
- [35] Garlasu D. , Sandulescu V. , Halcu L. Neculoiu G., Grigoriu O. , M. Marinescu and Marinescu V., "A Big Data implementation based on Grid Computing," Sinala, 2013.
- [36] Mahamud K. and Ruhana K. , "Big Data Clustering Using Grid Computing and Ant-Based Algorithm," Sarawak Malaysia, 2013.
- [37] L. Ware, "wisegeek," 2016. [Online] Available: <http://www.wisegeek.com/what-is-a-knowledge-grid.htm>. [Accessed 14 12 2016].
- [38] Marcos D. Assunção, Rodrigo N Calheiros, Silvia Bianchi , Marco A.S Netto, Rajkumar Buyyab, "Big Data computing and clouds: Trends and future directions," *J. Parallel Distrib. Comput* vol. 79, p. 3–15, 2015.
- [39] Brereton P. , Kitchenham B., Budgen D. Turner M. and Khalil M., "A Lesson: from applying the systematic literature review process with in the software engineering domain," *The Journal of system and software*, vol. 80, pp. 571-583 2007.
- [40] Chen L., Shadbolt N.R. and Goble C.A. "A Semantic Web-Based Approach to Knowledge Management for Grid Applications," *IEEE Transactions on Knowledge and Data Engineering*, vol 17, no. 2, pp. 283-296, 2007.
- [41] Darr E.D., Kurtzberg T.R. , "An Investigation of Partner Similarity Dimensions on Knowledge Transfer," *Organizational Behavior and Human Decision Processes*, vol. 82, no. 1, p. 28-44, 2000.
- [42] Wamba S.F. , Gunasekaran A. , Akter S. Ji-fan Ren S. ,Dubey R. and Childe S.J. "Big data analytics and firm performance Effects of dynamic capabilities," *Journal of Business Research*, vol. 70, p. 356-365, 2017.
- [43] Esposito C., Ficco M. and Palmieri F. "On Reliability and Security in Knowledge Grids," Naples, 2014.
- [44] Casaló L.V., Flavián C. and Guinalú M. "Relationship quality, community promotion and brand loyalty in virtual," *International Journal of Information Management*, vol. 30, p. 357–367, 2010.
- [45] Kambatla K., Kollias G., Kumar V. and Grama A., "Trends in big data analytics," *J. Parallel Distrib. Comput*, vol. 74, p 2561–2573, 2014.
- [46] Ma Z.Z. , Huang Y. , Wu J., Dong W. and Liyun Qi, "What matters for knowledge sharing in collectivistic cultures? Empirical evidence from China," *Journal of Knowledge Management*, vol. 18, no 5, pp. 1004 - 1019, 2014.
- [47] Abdullah R.H., "Halal Knowledge Grid Model for Promoting Knowledge Sharing Among Their Community of Practice," in *International Conference on Emerging Trends in Engineering and Technology* London, 2014.
- [48] Zhuge H., "Mapping Big Data into Knowledge Space with Cognitive Cyber-Infrastructure," *arXiv:1507.06500*, pp. 1-59, 2015.
- [49] Zhuge H., "The Knowledge Grid and Its Methodology," Beijing, 2005.
- [50] Abdullah R.H., Eri Z. D. and Talib A. M. "A model of knowledge management system for facilitating knowledge as a

- service (KaaS) in cloud computing environment," in *Research and Innovation in Information System (ICRIIS)*, Kuala Lumpur, Malaysia, 2011.
- [51] Abdullah R. H. , Eri Z. D. and Talib A M., "A model of knowledge management system in managing knowledge of software testing environment," in *Software Engineering (MySEC)*, Joho Bahru, Malaysia, 2011.