ISSN: 2836-8495

Research Article

Journal of Current Trends in Computer Science Research

# Requirements Engineering Approaches for Big Data Project Development

**Belachew Regane[1*], Sarah Beecham[2], Norah Power[3] and Dagmawi Lemma[1]**

[1]*Department of Computer Science, Addis Ababa University, Addis Ababa, Ethiopia*

[2]*Lero, the Science Foundation Ireland Research Centre for Software, Department of Computer Science & Information Systems, University of Limerick, Ireland*

[3]*Department of Computer Science & Information Systems, University of Limerick, Ireland*

**[*]Corresponding Author**
Belachew Regane, Department of Computer Science, Addis Ababa University, Addis Ababa, Ethiopia.

**Submitted**: 2024, Jun 03; **Accepted**: 2024, Jun 27; **Published**: 2024, Jul 05

**Citation**: Regane, B., Beecham, S., Power, N., Lemma, D. (2024). Requirements Engineering Approaches for Big Data Project Development. *J Curr Trends Comp Sci Res, 3*(4), 01-12.

## Abstract

***Context:*** *Today's digital world with millions of users results in vast amounts of data. This 'big data', characterized according to its volume, variety, velocity, and veracity, is impacting the lives of data users worldwide in many ways and has become important for day-to-day decision-making.*

***Problem:*** *Requirements engineering (RE) - approaches used to engineer user requirements - plays an important role in software development in general. However, when it comes to big data applications, it is unclear which requirements engineering approaches apply. There is therefore a need to investigate this further.*

***Objective:*** *This study aims to answer the following research questions: (1) How are requirements engineering (RE) activities performed to address the needs of stakeholders in the context of big data? (2) How are users' perspectives addressed in the RE activities in a big data context? (3) What are the requirements engineering approaches that have been proposed for big data project development?*

***Method:*** *To address our three research questions, we conducted a systematic mapping study focusing on requirements engineering and the existing requirements engineering approaches in software engineering in the context of big data projects.*

***Findings:*** *A total of 787 papers were examined, with 720 papers found through string-based search and a further 67 through snowball search. From the total search results, 17 relevant papers were identified and reviewed by applying inclusion-exclusion criteria. Findings show that in the realm of Requirements Engineering (RE) activities, there is a notable lack of emphasis on requirements negotiation, validation, and prioritization. Additionally, there is a scarcity of knowledge, methods, techniques, and tools tailored for conducting requirements engineering within the realm of big data. The user's role and perspective in RE are insufficiently considered. Although the goal-oriented RE approach is somewhat acknowledged among proposed methods, it has drawbacks such as neglecting the user's viewpoint, being relatively static and general in requirement representation, struggling to adapt to changing requirements, and having its effectiveness as the primary RE approach questioned. This approach primarily focuses on addressing the 'why' aspect of the system rather than the 'how' which aids in decision-making.*

***Conclusion:*** *Based on the findings, it is clear that there is a need for more research to be conducted to find a better way to have a suitable RE approach for big data application development.*

**Keywords:** Requirements, Requirements Engineering, Requirements Engineering Approach, Big Data, Big Data Application Development

## 1. Introduction

Big data technology benefits the user of data by providing insight for decision-making. However, big data application development has become a challenging task. The frequently changing requirements due to its Vs. characteristics (volume, variety, velocity, veracity) are among the factors that make the task challenging. Requirements engineering impacts positively or negatively the success of software development [1,2]. To have good requirements, a RE approach plays an important role in enhancing the software development process. Yet, according to Kourla et al [3]. traditional requirements engineering approaches cannot handle the frequently changing requirements of big data; and it is important to understand the challenges in RE in the case of developing big data technology. Thus, conducting a systematic mapping study is vital to get evidence of the challenges and potential opportunities by investigating previous works. This has been done by considering the state-of-the-art RE of the big data applications development process.

The main goal of systematic mapping studies is to get a comprehensive overview of a research area and identify the quantity and type of research and results available within the study area. Furthermore, it explained that the analysis of the results focuses on presenting the frequencies of publications for each category and this makes it possible to see which categories have been emphasized in past research and thus to identify gaps and possibilities for future research [4]. Similarly, it is discussed in that the rationale for conducting a systematic mapping is (i) to summarize the evidence for existing works in software engineering (ii) to identify gaps in existing research works (iii) to provide background for new research studies [5]. Therefore, the rationale for this systematic mapping is to get empirical evidence of the challenges and benefits of RE approaches that have been used in big data project development. According to Peterson et al [4]. findings, a systematic mapping study is different from a systematic review in terms of its goals, breadth, and depth. But they can complement each other. Therefore, "a systematic mapping can be conducted first, to get an overview of the topic area, and then the state of evidence in specific topics can be investigated using a systematic review" [4].

The result of this systematic mapping aims to benefit stakeholders in the domain area by providing evidence that justifies how to choose the right RE approach. Therefore, it could help software developers and requirements engineers to understand the advantages and disadvantages of those RE approaches. This enables them to choose the right approach to identify and analyze requirements for designing software. Thus, conducting a systematic mapping of RE approaches gives evidence of the challenges and potential benefits, investigating previous works by looking at the state-of-the-art in the area.

The rest of this technical report is structured as follows: Section two describes the method used for the study. Section three presents the results of the study. Section four discusses the findings in detail. Finally, section five concludes our findings and proposes future work.

## 2. Method

To address our research questions, we adopt Petersen et al.'s guidelines for conducting systematic mapping studies in software engineering, as outlined in the following sections [4].

### 2.1 Research Questions

We look to the literature to answer the following research questions:
1. How are requirements engineering (RE) activities performed to address the needs of stakeholders in the context of big data?
2. How are users' perspectives addressed in the RE activities in a big data context?
3. What are the requirements engineering approaches that have been proposed for big data project development?

### 2.2 Search Strategy

The strategy used for identifying search terms from Beecham et al. [6] was followed as a search strategy. Accordingly, major terms are derived from the research questions by identifying the population, intervention, and outcome; alternative spellings and synonyms for major terms are identified; keywords in any relevant papers we already have are checked; when the database allows, the Boolean OR is used to incorporate alternative spellings and synonyms; and the Boolean AND is used to link the major terms from the population, intervention, and outcome. Following these steps, the following search terms and strings are identified.

### 2.3 Constructing Search Terms

In this systematic mapping, the intervention is the requirements engineering approach, the population is stakeholders (users, software developers, and requirements engineers) and the outcome is requirements that satisfy the stakeholders' need to enhance software development of big data applications. This helped us as a basis to construct search terms while preparing a systematic mapping study protocol.

### 2.4 Search Terms and Synonyms

1. We identified search terms directly from our research questions, (e.g. Big Data, Requirements Engineering).
2: Synonyms were then identified for each of the search terms.
3 We grouped each search term and its synonyms (using operand OR).

### 2.5 Search Strings

Each group of search terms was combined using the 'AND' operand. As follows:
i. (("Big data") AND ("Requirements Engineering" OR "Requirements Analysis") AND (Stakeholder* OR User* ))
ii. (("Big data") AND ("Software Development" OR "Application Development" OR "Project Development") AND ("Requirements Engineering" OR "Requirements Analysis") AND (Approach* OR Method OR Process))

### 2.6 Resources Searched

The search strings shown under section 2.5 were used for both IEEE Xplore and ACM Digital Library. The search string (i) is used for research questions one and two. Search string (ii) is used for research question three. The dates of these searches were from 22 April 2021 to 30 May 2021.

In addition, we conducted a snowball search in which we looked at the references in the papers to identify papers of potential interest that were not found in our search terms. A snowball search helps to collect sufficient relevant documents for the study. Kitchenham and Charters recommend searching manually for relevant documents from "reference lists of relevant primary studies and review articles, grey literature (i.e. technical reports, work in progress)"[5].

### 2.7 Document Selection Criteria

#### 2.7.1 Inclusion and Exclusion Criteria

This section lists the inclusion and exclusion criteria that are used for considering or rejecting a published work as a form of evidence for addressing the research questions.

To be included, the study needs to comply with one or more of the following criteria:

• Subject matter criteria: Papers must

o Directly answer one or more research questions

o Focus on big data

o Focus on challenges in the context of requirements engineering

o Focus on requirements engineering activities in the context of user perspective

o Focus on solutions/approaches in the context of requirements engineering

Publication criteria: Papers must

o Be published as a journal article, conference/workshop proceedings

▪ The grey literature is included, such as articles published online as work in progress, reports [annual, research, technical, project, etc.], working papers, government documents, white papers, and evaluations. Such non-peer-reviewed articles are included to capture the views of a wider range of stakeholders.

o Be published from 2011 to 2021 (research has progressed over the past decade, so the focus will be on more recent work – since the literature review process included snowballing, key works published before this cut-off date of 2010 will be identified in the next phase);

o Report primary research (empirical studies)

▪ Literature surveys are included as an exception to the primary research but are used as background only, or for snowballing, and finding further empirical studies. Only empirical studies are used in our analysis.

o Be published in the English language

#### 2.7.2 Exclusion Criteria

Studies that fall into the following categories are excluded:

• Studies on the following topics are excluded:

o Studies on design;

o Studies on architecture;

o Studies on computer hardware and infrastructure, e.g. networks.

o Studies that do not explicitly discuss big data;

o Studies that do not focus on software engineering or requirements in any form (analysis, engineering).

• Types of publication excluded, include:

o Posters, opinion pieces papers, viewpoints, PowerPoint presentations, introduction to conferences;

o Books, book chapters, thesis, and dissertation;

o Keynote speech;

o Short papers (less than four pages);

o Summaries of conferences/editorials or guidelines/templates for conducting mapping studies;

o Studies are not accessible in full text.

#### 2.7.3 Selection Process

The selection process started by looking at, the titles of the journal articles, conference papers, or workshop papers. This is to check that they are relevant to the domain area. Following this, the abstracts part of the documents read to see their relevance. Then, if it is relevant, the entire document is reviewed. Then, three of the team members in the team validated each step of the sifting process separately.

The final set of papers selected is based on, i) the initial number of papers found (from running searches - search terms in DBs); ii) removal of duplicate papers; iii) checking titles, and abstracts; and iv) reading the full paper applying inclusion-exclusion criteria.

| **Process (paper acceptance based on inclusion/exclusion criteria)** | **Number of papers (Researcher 1)** | **Validation (Researchers 2 & 3)** |
|---|---|---|
| Starting set of from Electronic DB and Snowballing | 787 | Search strings validated Snowballing refs checked |
| Reading Title (removed 523) | 264 | 10% of 264 papers checked |
| Reading Abstract (removed 171) | 93 | 20% of 93 papers checked |
| Reading Full paper (removed 33) | 51 | 50% of 51 papers checked |
| Final Set of papers | 17 | 100% of accepted papers checked and coded |

**Table 1: Result of the Searching and Snowballing**

### 2.8 Data Synthesis

The data synthesis is a stage in the mapping process where extracted data (findings of individual studies) are combined and evaluated. It is a descriptive/narrative way to show the link between the intervention, population, context, sample sizes, outcomes, and study quality with the study questions. This helps to show how the data answers the research questions succinctly. The reason is the nature of the software engineering survey is qualitative–descriptive [5,7].

The data synthesis involved the first author extracting the data from each accepted primary study [5]. The extracted data was sent to three of the authors in this report, who validated the selection, querying some exclusions and inclusions (resulting in some additional inclusion and exclusion criteria, and changes in included papers). Each of the three researchers entered how they thought each accepted paper addressed the research question. Categorization of the data was done based on similarities/commonalities across all three researchers. The data were further categorized into classes to identify trends and topics. For example: synthesizing the many different types of activities and approaches, created classes 'RE activity types', and, 'RE approaches', respectively.

## 3. Findings

In this section, the results of the systematic mapping study are presented based on the analysis of the information extracted from the relevant papers selected, and organized by research questions.

RQ#1. ***How are requirements engineering activities performed to address the needs of stakeholders in the context of big data?*** Concerning the definition of big data, V-characteristics of big data are preferred in almost all the review papers. From these v-characteristics, volume, variety, velocity, and veracity are together used to define big data in five reviewed papers better than the other combination of Vs. On the other hand, five papers did not specify what type of definition they used. These results are shown in Table 2.

| V-characteristics used to define big data | No. of Vs used to define big data | No. of papers using the definition | References |
|---|---|---|---|
| Volume, variety, velocity, and veracity | 4 | 5 | [8]–[12] |
| Papers that have none of the Vs | 0 | 4 | [13]–[16] |
| Volume, variety, velocity, veracity, and value | 5 | 2 | [17], [18] |
| Volume, variety, velocity, and variability | 4 | 1 | [19] |
| Volume, Velocity, Variety, Variability, and Veracity | 5 | 1 | [20] |
| Volume, velocity, variety, variability, veracity, and visualization | 6 | 1 | [21] |
| Volume, velocity, variety, volatility, and variability | 5 | 1 | [3] |
| Volume, velocity, variety, consistency, variability, and volatility | 6 | 1 | [22] |
| Velocity, Variety, and Veracity. | 3 | 1 | [23] |

**Table 2: The Context of Big Data in Terms of V-Characteristics**

Figure 1 shows the number of papers published by several countries. The countries are Brazil, USA, Canada, Saudi Arabia, China, Germany and Indonesia. As shown below in the figure, RE papers in the context of big data are mainly published in the USA and Canada.

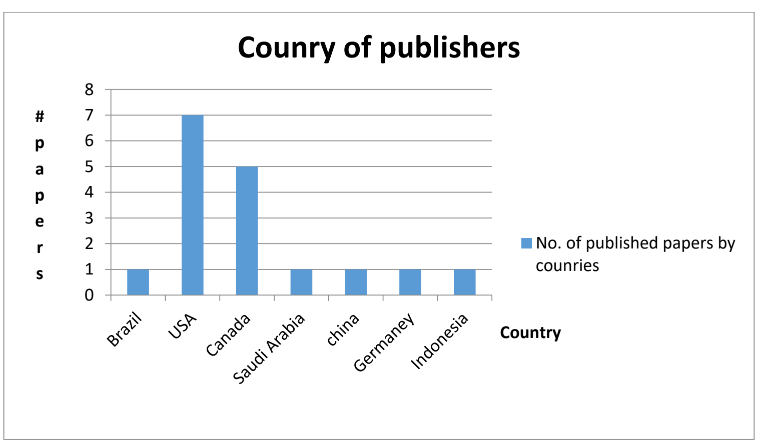


**Figure 1:** Number of RE Papers Published by Country

When we looked at the number of publications between 2011 and 2021 in which RE papers were published in the context of big data, 2017 was the year in which many papers were published. In these years five papers were published as shown in Figure 2 below.

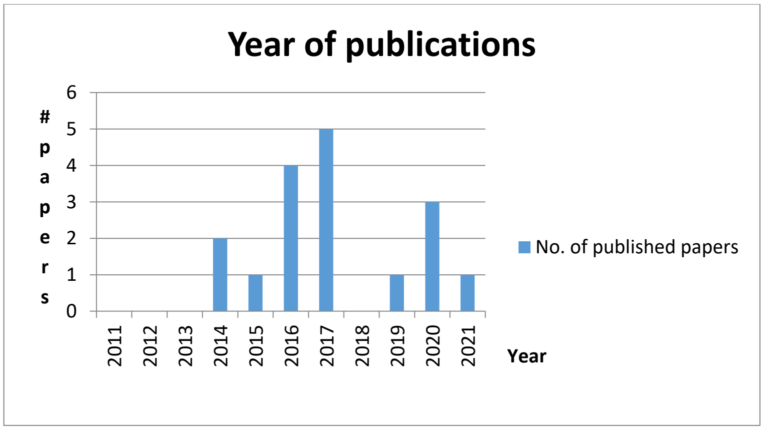


**Figure 2:** Numbers of RE Papers Published Between 2011 and 2021

To know how RE activities are performed to address the needs of stakeholders in the context of big data, 17 relevant papers were reviewed. From these papers, 14 papers were found discussing requirements engineering activities and challenges related to RE activities in the context of big data. Of 14 papers, 2 of them show requirements engineering activities in the context of user perspective. These are depicted in the following Table 3.

| Type of RE activities | Description | Count | References |
|---|---|---|---|
| All activities | The studies addressed all RE activities by proposing and building a model | 4 | [3], [10], [13], [14] |
| Elicitation | Studies those identify and gather requirements by using frameworks | 3 | [16], [19], [21] |
| Analysis | Studies that discovered the expectations of stakeholders using analysis techniques | 2 | [9], [22] |
| Specification | Studies those converted requirements into some standard form of users and systems requirements using the framework | 2 | [11], [17] |
| Management | A study that addresses understanding and controlling changes to system requirements by documenting requirements using a combination of data mining and semantic approaches | 1 | [16] |

**Table 3: RE Activities**

Concerning requirements engineering challenges in the context of big data, while reviewing the relevant papers to answer research question one, two papers were found discussing requirements engineering challenges and from these three papers, none of them discussed the challenges in the context of user perspective as depicted in Table 4.

| Type of challenges | Description | References |
|---|---|---|
| RE knowledge | There is a lack of sufficient knowledge, methods, techniques, and tools to do RE in the context of big data | [8], [20] |
| Big data characteristics in RE | The traditional approach couldn't handle integrating big data characteristics into the requirements | [8], [20] |

**Table 4: Challenges of RE in the Context of Big Data**

Table 4 shows that the research conducted by Arruda, and Laigner and Davoudian and Liu is open issues that need further investigation to have a piece of empirical evidence for understanding how to solve RE challenges [8,20]. Therefore, based on this gap analysis, we can conclude that RE needs further attention to have solutions as it will impact positively the success of big data project development.

RQ#2. ***How are users' perspectives addressed in the RE activities in a big data context?***

While searching and reviewing for relevant papers to answer research question two, only two papers were found discussing requirements engineering activities, and among those, no paper discussed open issues in requirements engineering in the context of user perspective. In all types of requirements engineering activities, and in elicitation and management of requirements engineering activities in the context of user perspective are discussed [18,16]. The number of papers that covered the types of requirements engineering activities from the users' perspectives is very few so it is difficult to understand how users' perspectives are addressed in the RE activities in a big data context.

RQ#3. ***What are the requirements engineering approaches that have been proposed for performing RE activities?***

In reviewing relevant papers to answer this research question, 14 of 17 reviewed papers were found discussing proposed RE approaches and contributions of the research results. The types of these research contributions and proposed approaches are models, frameworks, methods, and process diagrams, as depicted in Table 5, and GORE, Use case, and Task-directed RE approach as shown in Table 6.

| Types | Description | Count | References |
|---|---|---|---|
| Model | The study proposes and built ways of abstracting requirements to solve RE activities problems in the context of big data | 6 | [10], [13], [14], [18], [22], [23] |
| Framework | The study describes a proposal to solve problems of big data using RE frameworks | 4 | [3], [16], [17], [21] |
| Method | A study that describes procedures to be followed to solve problems of big data through RE activities | 3 | [9], [11], [15] |
| Process diagram | A study that describes how to process or processes of quality requirements can be addressed in the big data context | 1 | [12] |

**Table 5: Types of Contributions**

The above Table 5 provides a breakdown of different types of contributions related to big data and RE. It categorizes the contributions into four types: Model, Framework, Method, and Process diagram. Each type is described based on the study's focus and approach in addressing problems related to big data within the context of requirements engineering. The table also includes the count of each type of contribution and references to the corresponding studies.

| Classification | Description | Count | References |
|---|---|---|---|
| GORE Approach | The approach is used to build models and to address challenges faced in industries | 3 | [10], [15], [18] |
| Use case | The approach is used as a tool to extract requirements | 1 | [21] |
| Task-directed | This approach is used to address user perspective requirements elicitation & management issues | 1 | [16] |

**Table 6: RE Approaches**

Concerning RE approaches in the context of big data, five papers were found discussing RE approaches used to engineer requirements/proposed to be used to engineer requirements in the context of big data. These approaches are classified into three as depicted in Table 6.

## 4. Discussion

In this section, how the research questions are answered is presented. Based on the results of the mapping study, requirements engineering activities, and the challenges are presented focusing on the first research question. Then, user perspective issues are presented focusing on the second research question. Finally, the presentation emphasizes RE approaches to see how the third research question is answered. This has been done by discussing the contributions of the papers and approaches proposed/contributed to engineer requirements.

### 4.1 Requirements Engineering Activities in the Context of Big Data

#### 4.1.1 Studies that Address all RE Activities

Among the studies conducted to address the challenges of RE, Arruda's and Madhavji's work was reviewed [13]. The objective of the study was to get insight into different artifacts and their relationships in the context of big data project development for end-users focusing on RE. Based on the insight gained from the work, it is noted that they proposed a RE artifact model for Big Data end-user applications (BD-REAM). The proposed model is assumed to be used to perform all RE activities so that it gives room to "deal with what the customer wants, and how the system should behave during usage"[13]. The proposed model is built and validated in the industry by [14]. However, the model is built of many elements so that is complex and difficult to be used by the industry. Besides, the approaches they used to engineer the requirements are debatable so they might not be fit to handle the complex and frequently changing requirements of big data [24].

Another study that discussed generic RE activities was conducted by Coda et. al [18]. The purpose was to achieve the characterizing of requirements for the big data systems in Industry 4.0 – (smart manufacturing and the creation of intelligent factories), based on viewpoints, objectives, and expectations of stakeholders for its uses. They developed a model that reflects the viewpoints of stakeholders. Similarly, to handle problems arising from big data characteristics in requirements, Eridaputra et. al developed a generic requirement model for big data applications [10]. The making process of a generic model starts with deciding the general requirements of a big data application. A general requirement for big data applications is obtained from big data characteristics and challenges. This general requirement is then modeled using i* and KAOS. i. e the GORE approach. A study conducted by Kourla et al [3]. Proposed a generic requirements engineering framework, called REBD, in the context of big data. The framework was proposed by considering the limitations of traditional RE methods. However, it is a conceptual framework that is not validated empirically in the industry. Others used the GORE approach that can't handle the requirements of big data [24,25].

### 4.1.2 Requirements Elicitation

Concerning requirements elicitation, Al-Jaroodi and Mohamed discussed the V-characteristic of big data, emphasizing velocity focusing on real-time data correlating with the time constraint in a big data application [19]. The paper lacks scientific methods so it can be taken as a recommendation. Another study was conducted by Fox and Chang to have a consensus set of big data requirements across all stakeholders [21]. The data was collected using a use case as a technique. Then, by extracting requirements from the use case a framework of 35 generic requirements summarizing 439 specific requirements from the 51 use cases was developed. Wang et. al also developed a task-directed elicitation framework to elicit user requirements for big data systems by combining data mining with semantic approaches [16]. Their work focused on investigating the role of the user in requirements acquisition. However, this work has a limitation that arises from the limitation of data mining technology. Besides, the researchers are only concerned with requirements elicitation activities so the framework cannot address all the RE challenges in big data application development.

### 4.1.3 Requirements Analysis

Another study about requirements analysis conducted by Chen et. al claims that their proposed solution - Eco-ARCH addresses four big data value discovery requirements: (i) design thinking for innovation (ii) design for the open world (iii) integrating value discovery with value realization (iv) support for value experimentation and verification resulting in a new approach to eliciting requirements for, envisioning, and designing, systems in to compete in a fast-moving, open marketplace. Kozmina et al [9]. After their systematic survey, proposed a solution of requirements analysis that performs requirements at two phases in the development of the big data application [26]. (i) Before the development - it is possible to identify the user expectations timely and compare them with available resources, source data, quality, granularity, and available resources of the project i.e. budget, time, skills, and environment. (ii) When the data has already been loaded from the source systems - a user may not be able to define all information requirements. However, when all the data is collected, new relationships and values can be explored.

To identify, analyze, and determine requirements concerning requirements analysis Volk et al. Built an artifact model, using the design science approach. The contribution is an artifact in the form of a process model that uses the compound requirements, a big data classification framework, and Knowledge Discovery in Databases (KDD) [22]. The approach can help practitioners to be able to determine and formulate the requirements of big data projects.

### 4.1.4 Requirements Specification

To improve the process of big data collection, the study conducted by Al-Najran and Dahanayake developed a scenario-based big data collection framework [17]. The framework was intended to capture big data at the requirements engineering phase. The study aimed at accelerating“ the analysis time through data reduction by focusing on retrieving data from the source that meets the scenario”. This is because of the huge volume, velocity, and variety of big data; it is challenging to reduce the amount of data to be collected. Therefore, they developed the framework to solve the problems in the area of requirements specification activities.

To better understand quality requirements for big data systems and to propose an approach Noorwali et al. conducted research [11]. The proposed approach “consists of intersecting a big data characteristic with a quality attribute and then identifying the system's quality requirements that apply to that intersection. This ensures that big data characteristics are adequately addressed in the quality requirements specification.” However, the proposed approach is not empirically validated in the industry.

### 4.1.5 Requirements Management

Concerning requirements management activities, one of the contributions of research conducted by Wang et al is a task-directed framework developed to achieve a dynamic and easy-to-maintain requirement repository to manage the requirements of the big data system [16].

Finally, in requirements engineering activities in the context of big data through this mapping study, we observed that other requirements engineering activities like requirements negotiation, requirements validation, and prioritization are among requirements engineering activities that didn't get attention/were not discussed in the studies.

### 4.2 User Perspective in Requirements Engineering

In our study, we observed that there is no sufficient number of papers for review to see the status of user requirements in requirements engineering activities in the context of big data.

Therefore, it is difficult to conclude the user perspective of requirements engineering. In general, we can conclude that requirements engineering in the context of user perspective has not received attention in the context of big data so there is a need for further investigation.

### 4.3 Requirements Engineering Approaches and Proposed Solutions

#### 4.3.1 Goal-Oriented Requirement Engineering (GORE) Approach

Researchers conducted RE research in the context of big data and proposed a few approaches. GORE is among the proposed RE approach. GORE is an approach that answers the "why" questions on top of the waterfall approach which focuses on "what" questions of systems or business organizations. This gives a rationale for the functionality of the systems to help decisions [27].

To address the challenges industries face in their business, Supakkul and Chung proposed a goal-oriented modeling approach (GOMA) to big data analytics. The proposed approach - GOMA i) capturing business goals ii) understanding business problems iii) identifying business actions, and iv) taking business actions [15]. Using a goal-oriented approach Eridaputra et al. also developed a generic model to perform requirements engineering activities [10]. Coda et al. also used the goal-oriented requirement engineering (GORE) approach to develop a model that reflects stakeholders' requirements in Industry 4.0. However, in this approach, it is not clear how it handles the frequently changing requirements in new data [18]. It couldn't answer questions about how to handle issues related to the dynamicity of data and requirements for newly emerging data. Thus, the "how" question is an open issue in this approach. Besides, it doesn't answer questions relating to characteristics of big data – Vs (volume, variety, velocity, veracity) for real-time data. In addition, the goal-oriented approach is on the debate as an approach to be used for RE. Because of, according to Yu & Mylopoulos "In some RE frameworks, they are central; in others, they play a supporting role [24]." The goal is not stable. It will change from time to time. However, it takes assumption for stability [25].

According to Aljahdali et al., the goal-oriented approach is an activity that is performed in the early phase of RE [27]. This means it supports activities that have been done before the formulation of initial requirements. Since the requirements are identified at the very beginning, the approach has no space for iteration to handle changes that arise from customers. Arruda et al., also spell out that their model built based on a goal-oriented approach is not generic so it can't solve all the problems related to big data [14]. Therefore, as reported by Horkoff et al. "goal-oriented approach is preferable in sub-RE like in agent orientation, aspect orientation, business intelligence, model-driven development, and security" than for big data software development [28].

Using the GORE approach, models are built and proposed as a solution for big data project development. For example, Arruda and Madhavji proposed an RE Artefact model that could enable the developers of big data software to "deal with what the customer wants, and how the system should behave during usage [13]." This model is built and evaluated in the industry and modified by Arruda et al [14]. Eridaputra et al. also developed a generic requirements model using i* and KAOS. i. e GORE approach for big data application [10].

A model is a representation of a real-world problem in easy ways so that people can understand it. A model can be built as a solution to solve problems of requirements engineering in the context of big data application development. Al-Jaroodi and Mohamed recommended an agile process development model to develop big data application software successfully [19]. Coda et al. proposed a model to develop " the system that reflects stakeholder requirements for big data systems" [18]. Volk et al. developed a process model which is in the form of an artifact using the design science approach, using the compound requirements, from a big data classification framework, and Knowledge Discovery in Databases [22]. However, these requirements engineering models can't handle the requirements of big data. This is due to the Vs characteristics of big data.

#### 4.3.2 Task-Directed Approach

Task-directed approach is another approach proposed as a solution for big data challenges. According to Christian and Paech, Task-oriented RE (TORE) emphasizes a comprehensive RE specification that covers stakeholders' tasks, data, system functions, interactions, and UI. Task orientation focuses on the RE process and helps to deliver software that satisfies user needs [29].

In the above context, to investigate the importance of the role of stakeholders, in users perspective, while building practical system functions in terms of requirement acquisition in data analytics or management, Wang et al. conducted a research study and proposed a task-directed approach for requirements elicitation and management [16]. The proposed approach takes semantics extracted resources from online as an important part of domain knowledge. "This involves ontology construction from task-related documents and enrichment with external ontology repositories or linked open data. The interaction of users with the system when performing tasks is also regarded as contextual semantics to refine the requirement with the evolution of user interest." However, task-directed requirements engineering has the limitations of (i) it is only about elicitation management and (ii) its limitation arises from the limitation of data mining and semantic technology as well as from data quality. So, in this research study context, the finding and proposed solution can't solve the challenges of big data requirements which frequently have been changing due to its characteristics.

One of the contributions proposed using TORE is the framework. A framework is a description of guidelines to solve problems. There are some RE frameworks proposed and developed to solve problems in big data applications. For example, the contribution of is a TAsk-Directed ELicitation (TADEL) framework which is a user-centric framework to handle the changing needs of users [16].

However, this requirements-related proposed solution is specific to one of the requirements engineering activities so it couldn't solve all the challenges of big data application development. Another framework is the contribution of Al-Najran and Dahanayake, who developed " a scenario-based big data collection framework that performs" requirements specification activities. Another framework contributed by Kourla et al. is a big data requirements engineering framework named REBD [17,3]. The frameworks developed as it can "first identify the project type and then perform corresponding RE processes" [3].

#### 4.3.3 Use Case

Use case is another approach used to develop a framework for big data solutions. Use cases are a requirements discovery technique that was first introduced in the objectory method. They have now become a fundamental feature of the unified modeling language (UML) [30]. For example, to have common big data requirements that can be agreed on by all stakeholders, Fox and Chang conducted research and developed a framework of requirements [21]. It is to have "a consensus set of big data requirements across all stakeholders". To develop the framework the researchers used a use case as a requirements elicitation technique. The framework is vendor-neutral and technology-agnostic. According to Jacobson use cases and use case-based requirements elicitation is now widely used in requirements engineering activities since UML is a de facto standard for object-oriented modeling [30]. Use cases are effective techniques for eliciting requirements from stakeholders who interact directly with the system. Each type of interaction can be represented as a use case. However, because they focus on interactions with the system, they are not as effective for eliciting constraints or high-level business and non-functional requirements or for discovering domain requirements.

To solve the problem of requirements engineering, a few researchers proposed methods as a contribution. The method is the steps or procedures that should be followed by the researchers to achieve the stated objectives of the study [31]. Chen et al in the study try to address the issue of requirement analysis using the value-based requirements analysis method [9]. With the objective "to better understand quality requirements for big data systems", Noorwal et al. conducted a research study and came up with a new method [11]. This method "consists of intersecting a big data characteristic with a quality attribute and then identifying the system's quality requirements that apply to that intersection" [11].

Another contribution proposed as a solution for the big data challenge is the process diagram. A process diagram is a visual that illustrates a set of steps and decision points for executing specific tasks. In this context, to introduce quality requirements of security and performance, early in the software development lifecycle, Sachdeva and Chung developed a process diagram [12]. The process diagram is developed using agile methodology. A case study was used to validate it in the industry. Then, based on validation results, they modified it for better use. However, the study didn't say something about big data specifically about big data characteristics and how to integrate these characteristics into quality requirements while engineering the requirements, in designing and developing big data application software. In addition, it considered only quality requirements of security and performance so it is difficult to generalize based on these two quality attributes.

### 4.4 Implications of Findings

This subsection presents the implications of the findings - that is how RE performed to satisfy stakeholders' needs while developing big data applications.

#### A. Concerning RE Activities (RQ#1):

1. Requirements negotiation, validation, and prioritization do not get the attention.
2. There is a lack of sufficient knowledge, methods, techniques, and tools to do requirements engineering in the context of big data.

#### B. Concerning User Perspective RE Activities (RQ#2):

The big data applications are developed for users/end-users so that their needs should be given attention to satisfy users. However, according to our observations:

1. User role and RE from the user perspective do not receive attention.
2. One of the studies that proposed a task-directed approach - is an approach that focuses on the user perspective has the following limitations:

a. It is only about elicitation management

b. Its limitation arises from the limitation of data mining and semantic technology as well as from data quality so that can't address the challenges of big data in terms of big data characteristics.

#### C. Concerning RE Approaches (RQ#3):

1. From the proposed RE approaches, the goal-oriented RE approach gets some attention. However, the approach has the following limitations:

a) It does not address the user's perspective.

b) It is relatively static and general to represent requirements.

c) It cannot handle changing requirements.

d) The GORE approach received more attention than other approaches used in the reviewed paper. However, it is debatable whether it can be used as the main RE approach.

e) It focuses on the 'why' issue of the system and then 'how' – that can assist decision-making.

2. The agile requirements engineering approach is recommended as a potential approach for changing requirements. So, agile has the potential to be used as a method to solve the issues in RE for big data context. Because it can handle the changing requirements of big data.

#### D. General Analysis of Existing Works:

Existing works have the following limitations:-

1. the contributions are developed of many elements so that complex, couldn't serve at a domain-specific level so impractical at industries [32].
2. can't handle the frequently changing requirements of big data; the solutions are not dynamic or agile

3. focus on one aspect; however, big data challenges need to integrate different aspects like V-characteristics, quality requirements, user requirements, agility concepts, etc.
4. the solutions couldn't address all stakeholders' needs sufficiently, E.g. the users' needs/perspective.
5. missed guidelines for choosing the right knowledge to enhance collaboration among all stakeholders involved in the development process [8,33].

### 4.5 Concluding Remarks

The implications of the systematic mapping study for big data application development are presented in this section.

To answer the research questions of the study, 720 papers were found through string-based search and 67 further papers through snowball search; in total 787 were found. After applying the inclusion-exclusion criteria and reviewing the selected papers, 17 relevant papers were identified.

The study found that generally speaking, concerning RE activities, very few studies were conducted that could be applied to engineer requirements and develop big data applications. When we looked at how requirements engineering activities were performed in addressing the needs of stakeholders in the context of big data, we observed that few studies tried to address these issues. Four studies tried to come up with solutions that can address all RE activities whereas other studies tried to address single or several RE activities solutions in their study. However, certain RE activities like requirements validation, negotiation, and requirements prioritization are not covered in the studies.

Furthermore, the challenges related to lack of RE know-how in the context of big data and how to integrate big data characteristics into quality requirements are the challenges discussed in the studies and are open to being investigated further. No particular solutions to these challenges were offered.

Concerning the data user perspective (RQ#2) this area did not get attention, indicating a potential area for future research.

Concerning RE approaches used (RQ#3), the goal-oriented requirements engineering approach is used in three studies and the task-directed approach is used in one study to address requirements engineering issues in the context of user perspective. The contributions of the studies include models, frameworks, and methods, the model and framework get better attention than the methods.

From the findings, we conclude that studies conducted about RE in the context of big data application development are few. Therefore, (i) there is a lack of any empirical study about RE approaches that are appropriate for requirements engineering activities to be used for big data applications development. (ii) there is a lack of evidence for usable frameworks, and models that can help to engineer requirements for big data applications development in industries. (iii) there is a lack of sufficient evidence that can support choosing the right approach to perform requirements engineering activities to enhance big data application development. Thus, it is clear that there is a need for more research to be conducted to find a better way to have a suitable RE approach for big data application development [34-50].

## References


1. Chen, F., Power, N., Collins, J. J., & Ishikawa, F. (2019, April). Contemporary requirements challenges and issues: an empirical study in 11 organizations. In *Proceedings of the 34th ACM/SIGAPP symposium on applied computing* (pp. 1592-1599).
2. Eberlein, A., & Leite, J. C. S. P. (2002, September). Agile requirements definition: A view from requirements engineering. In *Proceedings of the international workshop on time-constrained requirements engineering (TCRE'02)* (pp. 4-8).
3. Kourla, S. R., Putti, E., & Maleki, M. (2020). REBD: A Conceptual Framework for Big Data Requirements Engineering. *arXiv preprint arXiv:2006.11195*.
4. Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008, June). Systematic mapping studies in software engineering. In *12th international conference on evaluation and assessment in software engineering (EASE)*. BCS Learning & Development.
5. Keele, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* (Vol. 5). Technical report, ver. 2.3 ebse technical report. ebse.
6. Beecham, S., Baddoo, N., Hall, T., Robinson, H., & Sharp, H. (2006). Protocol for a systematic literature review of motivation in software engineering.
7. Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., & Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of systems and software, 80*(4), 571-583.
8. Arruda, D., & Laigner, R. (2020, December). Requirements Engineering Practices and Challenges in the Context of Big Data Software Development Projects: Early Insights from a Case Study. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 2012-2019). IEEE.
9. Chen, H. M., Kazman, R., Garbajosa, J., & Gonzalez, E. (2016, May). Toward big data value engineering for innovation. In *Proceedings of the 2nd International Workshop on BIG Data Software Engineering* (pp. 44-50).
10. Eridaputra, H., Hendradjaya, B., & Sunindyo, W. D. (2014, November). Modeling the requirements for big data application using goal oriented approach. In *2014 international conference on data and software engineering (ICODSE)* (pp. 1-6). IEEE.
11. Noorwali, I., Arruda, D., & Madhavji, N. H. (2016, May). Understanding quality requirements in the context of big data systems. In *Proceedings of the 2nd International Workshop on BIG Data Software Engineering* (pp. 76-79).
12. Sachdeva, V., & Chung, L. (2017, January). Handling non-functional requirements for big data and IOT projects in Scrum. In *2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence* (pp. 216-221). IEEE.

13. Arruda, D., & Madhavji, N. H. (2017, December). Towards a requirements engineering artefact model in the context of big data software development projects: Research in progress. In *2017 IEEE international conference on big data (big data)* (pp. 2314-2319). IEEE.
14. Arruda, D., Madhavji, N. H., & Noorwali, I. (2019, July). A Validation Study of a Requirements Engineering Artefact Model for Big Data Software Development Projects. In *ICSOFT* (pp. 106-116).
15. Supakkul, S., Zhao, L., & Chung, L. (2016, June). GOMA: supporting big data analytics with a goal-oriented approach. In *2016 IEEE International Congress on Big Data* (BigData Congress) (pp. 149-156). IEEE.
16. Wang, P., Tao, K., Gao, C., Ning, X., Gu, S., & Deng, B. (2017, November). Eliciting big data requirement from big data itself: A task-directed approach. In *2017 6th International Workshop on Software Mining (SoftwareMining)* (pp. 17-23). IEEE.
17. Al-Najran, N., & Dahanayake, A. (2015). A requirements specification framework for big data collection and capture. In *New Trends in Databases and Information Systems: ADBIS 2015 Short Papers and Workshops, BigDap, DCSA, GID, MEBIS, OAIS, SW4CH, WISARD, Poitiers, France, September 8-11, 2015. Proceedings* (pp. 12-19). Springer International Publishing.
18. Coda, F. A., de Salles, R. M., Junqueira, F., Santos Filho, D. J., Silva, J. R., & Miyagi, P. E. (2018, November). Big data systems requirements for Industry 4.0. In 2018 *13th IEEE International Conference on Industry Applications (INDUSCON)* (pp. 1230-1236). IEEE.
19. Al-Jaroodi, J., & Mohamed, N. (2016, November). Characteristics and requirements of big data analytics applications. In *2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC)* (pp. 426-432). IEEE.
20. Davoudian, A., & Liu, M. (2020). Big data systems: A software engineering perspective. *ACM Computing Surveys (CSUR), 53*(5), 1-39.
21. Fox, G., & Chang, W. (2014, March). Big data use cases and requirements. In *1st Big Data Interoperability Framework Workshop: Building Robust Big Data Ecosystem ISO/IEC JTC* (Vol. 1, pp. 18-21).
22. Volk, M., Jamous, N., & Turowski, K. (2017, August). Ask the Right Questions: Requirements Engineering for the Execution of Big Data Projects. In *AMCIS*.
23. Al-Jaroodi, J., Hollein, B., & Mohamed, N. (2017, January). Applying software engineering processes for big data analytics applications development. In *2017 IEEE 7th annual computing and communication workshop and conference (CCWC)* (pp. 1-7). IEEE.
24. Yu, E., & Mylopoulos, J. (1998, June). Why goal-oriented requirements engineering. In *Proceedings of the 4th international workshop on requirements engineering: Foundations of software quality* (Vol. 15, pp. 15-22).
25. Anton, A. I. (1996, April). Goal-based requirements analysis. In *Proceedings of the second international conference on requirements engineering* (pp. 136-144). IEEE.
26. Kozmina, N., Niedrite, L., & Zemnickis, J. (2019). Perspectives of information requirements analysis in big data projects. In *Databases and Information Systems X* (pp. 109-124). IOS Press.
27. S. Aljahdali, J. Bano, and N. Hundewale, "Goal Oriented Requirements Engineering -A Review Goal Oriented Requirements Engineering - A Review," Isca Caine, no. January 2011, pp. 328–333, 2015.
28. Horkoff, J., Aydemir, F. B., Cardoso, E., Li, T., Maté, A., Paja, E., ... & Giorgini, P. (2019). Goal-oriented requirements engineering: an extended systematic mapping study. *Requirements engineering, 24*, 133-160.
29. Kücherer, C., & Paech, B. (2017, April). A task-oriented requirements engineering method for personal decision support systems-a case study. In *International Conference on Enterprise Information Systems* (Vol. 2, pp. 99-110). SCITEPRESS.
30. Jacobson, "Use Cases -- Yesterday , Today , and Tomorrow Yesterday," vol. 3, no. 3, 2004.
31. Wahyuni, D. (2012). The research design maze: Understanding paradigms, cases, methods and methodologies. *Journal of applied management accounting research, 10*(1), 69-80.
32. Misra, S., Kumar, V., & Kumar, U. (2005, May). Goal-oriented or scenario-based requirements engineering technique-what should a practitioner select?. In *Canadian Conference on Electrical and Computer Engineering, 2005*. (pp. 2288-2292). IEEE.
33. Schön, E. M., Thomaschewski, J., & Escalona, M. J. (2017). Agile Requirements Engineering: A systematic literature review. *Computer standards & interfaces, 49*, 79-91.
34. Kourla, S. R., Putti, E., & Maleki, M. (2020). REBD: A Conceptual Framework for Big Data Requirements Engineering. *arXiv preprint arXiv:2006.11195.*
35. Arruda, D., & Laigner, R. (2020, December). Requirements Engineering Practices and Challenges in the Context of Big Data Software Development Projects: Early Insights from a Case Study. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 2012-2019). IEEE.
36. Chen, H. M., Kazman, R., Garbajosa, J., & Gonzalez, E. (2016, May). Toward big data value engineering for innovation. In *Proceedings of the 2nd International Workshop on BIG Data Software Engineering* (pp. 44-50).
37. Eridaputra, H., Hendradjaya, B., & Sunindyo, W. D. (2014, November). Modeling the requirements for big data application using goal oriented approach. In *2014 international conference on data and software engineering (ICODSE)* (pp. 1-6). IEEE.
38. Noorwali, I., Arruda, D., & Madhavji, N. H. (2016, May). Understanding quality requirements in the context of big data systems. In *Proceedings of the 2nd International Workshop on BIG Data Software Engineering* (pp. 76-79).
39. Sachdeva, V., & Chung, L. (2017, January). Handling non-functional requirements for big data and IOT projects in Scrum. In *2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence* (pp. 216-221). IEEE.

40. Arruda, D., & Madhavji, N. H. (2017, December). Towards a requirements engineering artefact model in the context of big data software development projects: Research in progress. In *2017 IEEE international conference on big data (big data)* (pp. 2314-2319). IEEE.
41. Arruda, D., Madhavji, N. H., & Noorwali, I. (2019, July). A Validation Study of a Requirements Engineering Artefact Model for Big Data Software Development Projects. In *ICSOFT* (pp. 106-116).
42. Supakkul, S., Zhao, L., & Chung, L. (2016, June). GOMA: supporting big data analytics with a goal-oriented approach. In *2016 IEEE International Congress on Big Data (BigData Congress)* (pp. 149-156). IEEE.
43. Wang, P., Tao, K., Gao, C., Ning, X., Gu, S., & Deng, B. (2017, November). Eliciting big data requirement from big data itself: A task-directed approach. In *2017 6th International Workshop on Software Mining (SoftwareMining)* (pp. 17-23). IEEE.
44. Al-Najran, N., & Dahanayake, A. (2015). A requirements specification framework for big data collection and capture. In *New Trends in Databases and Information Systems: ADBIS 2015 Short Papers and Workshops, BigDap, DCSA, GID, MEBIS, OAIS, SW4CH, WISARD, Poitiers, France, September 8-11, 2015. Proceedings* (pp. 12-19). Springer International Publishing.
45. Coda, F. A., de Salles, R. M., Junqueira, F., Santos Filho, D. J., Silva, J. R., & Miyagi, P. E. (2018, November). Big data systems requirements for Industry 4.0. In *2018 13th IEEE International Conference on Industry Applications (INDUSCON)* (pp. 1230-1236). IEEE.
46. Al-Jaroodi, J., & Mohamed, N. (2016, November). Characteristics and requirements of big data analytics applications. In *2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC)* (pp. 426-432). IEEE.
47. Davoudian, A., & Liu, M. (2020). Big data systems: A software engineering perspective. *ACM Computing Surveys (CSUR), 53*(5), 1-39.
48. Fox, G., & Chang, W. (2014, March). Big data use cases and requirements. In *1st Big Data Interoperability Framework Workshop: Building Robust Big Data Ecosystem ISO/IEC JTC* (Vol. 1, pp. 18-21).
49. Volk, M., Jamous, N., & Turowski, K. (2017, August). Ask the Right Questions: Requirements Engineering for the Execution of Big Data Projects. In *AMCIS*.
50. Al-Jaroodi, J., Hollein, B., & Mohamed, N. (2017, January). Applying software engineering processes for big data analytics applications development. In *2017 IEEE 7th annual computing and communication workshop and conference (CCWC)* (pp. 1-7). IEEE.