A Review for Analytics Requirements Processes: From Malaysia Public Sector Perspectives

Zuliana Daud[1] , Suraya Ya'acob[2] , Nur Azaliah Abu Bakar[2]
Surya Sumarni Hussein[3]
[1]*Institute Diplomacy and Foreign Relations, Kuala Lumpur.*
[2]*Faculty Technology and Informatics Razak, Universiti Teknologi Malaysia.*
[3]*Faculty of Computer and Mathematical Science, Universiti Teknologi Mara Shah Alam.*

zuliana@idfr.gov.my[1], suraya.yaacob@utm.my[2], azaliah@utm.my[2],suryasumarni@fskm.uitm.edu.my[3],

Article history
Received: 15 May 2022

Received in revised form:
01 June 2022

Accepted:
15 June 2022

Published online:
30 June 2022

*Corresponding author
suraya.yaacob@utm.my

Abstract

Big Data Analytic (BDA) projects in the Malaysia's public sector become challenging because it involves complex processes of diverse data sets among agencies, ministries and stakeholders to uncover valuable insights. One of the critical challenges is to ensure the analytics output is meaningful and valuable in facilitating business decisions in the public sector. Hence, the clarity for analytics requirements or typically known as Business Requirement Specification (BRS) is essential. Clear analytics requirements can help business and technical persons understand what to expect, deliver, and change during the BDA project development. Furthermore, it is important to understand the requirements of the business and align them to the analytics features. Currently, the level of alignment between business and technology is constantly uncertain and the process of developing analytics requirements mostly relies on current methodologies and processes which are still unsure of their ability to handle the analytics requirements. Hence, this paper intends to investigate further, explore and analyse the current methodologies and their suitability to handle analytics requirements. These findings can be extended to develop more practical processes in guiding the analytics requirements in the future.

Keywords— big data and analytics, analytics requirements, business requirement specification, methodology, process, public sector.

1. Introduction

Nowadays many organizations develop Big Data and Analytics (BDA) projects. Analytics provides organization with an advantage in that it improves their competitive positions. Gartner is forecasting that businesses that are not investing significantly in their analytics by the end of 2020 is not in business in 2021. Big data analytics is the vast and complex method of analysing large and varied data sets to discover knowledge. [1]. BDA projects are the often-complex processes of examining large and varied data sets or big data to uncover information, such as hidden patterns, unknown correlations, market trends, and customer preferences, which can help in decision making decisions. BDA projects have multiple layers of data and information which may all need to be detailed by requirements. These projects involve data itself, the data operation, data formatting, data distribution, and data visualization. In the past, organizations that deployed analytics projects primarily focused on descriptive analytics [2]. This means examining reports that inform stakeholders what is going on or has occurred in their organization. Recent developments suggest a move to use predictive and prescriptive analytics for more organizations. BDA projects will add business value, as it stimulates data-driven decision-making capabilities, in which case judgements are often more precise than when they are based solely on intuition or experience [3].

The BDA projects can be complicated since they are positioned at the intersection of different disciplines including data management, analytical processes and instruments, functional areas, forms and sources of data and the business processes that BDA projects help [4]. The multiple elements in BDA projects are interdependent and therefore need much synchronization to manage dependency among the project BRS. The main goal of business analytics initiatives is to build software systems that can transform massive, highly complex data sets into displays of meaningful information. The introduction of emerging technology requires organisations to understand. For most Information Communication and Technology (ICT) projects, reports represent just a small portion of the functionality implemented. But on business analytics initiatives, comprehensive reports, visuals, and users' ability to manage their contents identify the key capabilities.

2. Problem Background

The idea of getting results from Big Data and Analytics (BDA) projects is quickly becoming the norm in the Big Data era. With the emergence of Big Data, the concept of having a result from BDA projects is rapidly becoming the norm in the Big Data age. Today, the most innovative businesses leverage data power to expand their sectors into new industries. It helps to differentiate champions from underperformers and improve market growth, transformation, and innovation of creativity, development, and optimization through new technologies [5].

In Big Data projects, poor coordination and unclear objectives were described as key problems [6]. The level of alignment between technology and business is always uncertain in technology project management [7]. To manage BDA projects, most developers and organizations rely on general Information Communication and Technology (ICT) methodology or data science methodology such as CRISP-DM, SEMMA and SDLC that are yet unsure of their capacity to handle the alignment between technical and business from BDA projects perspectives.

Hence, the literature review found the importance of Business Requirement Specification (BRS) clarity to ensure the effective alignment. It can facilitate the business and technical people aware of what to expect, deliver and change during BDA projects development. The comprehensive BRS standard and methodology in project management is needed to reduce the failures of BDA projects. The standards and guideline would help control BDA projects, thereby helping to reduce project delays, thus on-time performance of projects within the current budget. Finally, BRS should facilitate in promoting coherent and transparency of expected outcomes and its changes due to the real condition of the BDA project's progress.

3. Literature Review for Business Requirement Specification

Analytics requirement typically known as a Business Requirement Specification in among the ICT Project in the Malaysia's Public Sector. Business requirements are a set of requirements that define the need for a solution, and the result that must be achieved by a solution [8]. The business requirements describe the business problems that can be solved by the proposed software [9]. Business requirement is a need from a business perspective, user, client, or employee to be analysed and processed for the system development and design. Business requirements also capture the business process or procedure or activities in order to achieve business goals. The scope of the solution is also specified by business requirements

A specification of requirements is a collectively defined set of requirements that fulfil certain criteria, usually for a system or component [10]. A significant number of

theoretical research have been performed and the most successful practices have been established [11]. The requirement specification specifies exactly how a software system or project should be implemented. Requirement specifications also define how the system or projects operate and what are the key performance that will give customer satisfaction.

To understand more on business requirement specification (BRS), the searching for relevant articles, using manual and electronic databases has been done. The timeframe of the research papers was limited during the search process between 2015 and 2020. By reading the title, abstract and keywords, the retrieved article was chosen, and irrelevant studies were omitted based only on BRS and requirement engineering. There eight (8) relevant articles have been found and summarized in Table 1.

Table 1. Summary of the Literature Review BRS for BDA

Study ID	Author(s)/ Organization	Title	The BRS for BDA mention in the study
R001	Hamza Hussein Altarturi et al (2017)	A Requirement Engineering Model for Big Data Software	• BRS should be data-centred rather than user-centred. • Project goals should be set early on. • New process: Data acquisition, Data analysis and value discovery, and Use cases consolidation. • Proposed a model not validate.
R002	Darlan Arruda (2017)	Requirements Engineering in the Context of Big Data Applications: Exploratory Research	• Difficulty seeking proper framework. • Need for addressing Big Data. • The lack of specific goals.
R003	Evellin Cristine Souza Cardoso et al (2014)	Requirements Engineering Based on Business Process Models: A Case Study	• Process models provided an effective and full specifications specification. • An approach is important for establishing unique specifications.
R004	Elsa Marcelino-Jesus et al (2014)	A Requirements Engineering Methodology for Technological Innovations Assessment	• Business Requirements goals, objectives, or needs of the enterprise. • Create approaches that can be used by other systems and applications.
R005	Ing. Arno Kühna et al (2018)	Analytic Canvas - A Framework for the Design and Specification of Data Analytics Projects	• Multiple stakeholders are involved. • Lack of cooperation. • Use CRISP-DM methodology. • Data science project framework. • Framework not validated.
R006	Jeffrey S. Saltz et al (2017)	Comparing Data Science Project Management Methodologies via a Controlled Experiment	• Test three (3) methodology Agile Scrum, Agile Kanban and CRISP-DM. • A better technique is required.
R007	Rafael Gorski Moreno Souza et al (2016)	Problem-Based Software Requirements Specification	• Human-centred activity. • Classify Customer Problem, Customer Need. • Software Glance concept (solution overview).
R008	Gerrit Muller (2020)	Requirements Capturing by the System Architect	• Multiple complementary viewpoints. • Requirements both top-down, as well as bottom-up. • The notion of "business key drivers" is introduced and a method is described to link these key drivers to the product specification

4. Methodology and Process Model Related to Analytics Requirements

In the Cambridge Dictionary, methodology is a set of ways to do, teach, or study something. In system development, methodologies are proposed to improve the management and control of the development process, to structure and simplify the process, and to standardise the development process and product by defining the activities to be performed and the techniques to be employed [12]. Meanwhile, process model is general approaches for organizing a project into activities. In general, process models are defined requirements and instructions for repeated tasks and processes that contain criteria, techniques, processes, and practices [13].

The graphical representation of business processes or workflows is known as process modelling, and it involves mapping out the entire process from start to finish [14]. Process models have been increasingly influential as a tool for organizational development in recent decades [15] and used in software development to manage cost, time, and quality concerns as well as changing client needs [16]. As for the output, process model creates a visual representation of development steps and processes, allowing for easier inspection and comprehension of how processes work in their current condition and how they might be improved. For further understanding, this research explores the current methodologies/process model related to gather analytics requirements. There are three perspectives which are related to Requirement Engineering, big data analytics and methodology that specifically develop and used by ICT Malaysian Public Sectors. Each paragraph below will discuss in detail about these.

4.1. Requirement Engineering.

Requirements engineering is about identifying, recording the requirements, and managing the requirements process. Understanding what the customer needs, review, feasibility assessment, and negotiation of a solution that is acceptable, simple identification of a solution, validation of specifications and management of requirements in a working system.

This model is a sequential model that the process is done after one another. Requirement's engineering is a structured and disciplined approach for the design and management of requirements [10]. RE process and methodology is a sequential model that the process is done after one another. RE process as follow:

Table 2. Requirement Engineering Process (IREB, 2015)

Process	Details
i. Feasibility Study	• To create the reasons for developing the software that is acceptable to
ii. Requirement Elicitation and Analysis	• The gathering of requirements, requirements are identified with the help of customers and existing systems processes, if available. • Requirements are analyzed to identify inconsistencies, defects, omission.
iii. Software Requirement Specification	• Document which is created by a software analyst after the requirements collected from the various sources. • To write the requirement in technical language so that they can be understood and beneficial by the development team.
iv. Software Requirement Validation	• After requirement specifications developed, the requirements discussed in this document are validated

From the understanding, the requirement engineering process did not cater anything related on data analytic. The requirement engineering model gather information about the business process and focus on application and software process to meet user needs. This process explains more on how to elicitate functional requirements and not so much about what and how to develop the analytics requirements.

4.2. Requirements Process in the Big Data Analytics

The research refers to two main processes related to the big data analytics methodologies which are CRISP-DM and TDSP. CRISP-DM approach offers an excellent plan for designing data mining projects. It is a well-researched process. In 2016, CRISP-DM has been improvised by IBM in term of project and communication management and been known as Analytics Solutions Unified Method for Data Mining (ASUM-DM). Figure 1 show the CRIPS-DM process and explanation.

This model is based on an idealised sequence of events. In reality, multiple tasks may be conducted in various orders, and it sometimes becomes necessary to replicate acts that have already been done. This process model offers a flexible structure that enables one to develop, design, construct, test, and deploy machine learning solutions. The activities in this model represent an ideal sequence. Most of the work can be done in a different order and it is usually essential to repeat tasks in order to return to earlier activities. The model doesn't attempt to include all conceivable data mining pathway.

There are five processes in the CRISP-DM which are business understanding, data understanding, data preparation, evaluation and deployement. The most critical process that related to requirement is the business understanding process. It aims to have a clear understanding of what the organization want to accomplish from a business perspective. Generally, four sub-processes are determine business objectives, assess situation, determine data mining goals and produce project plan.

Figure 1. ASUM-DM (Improvised version of CRIPS-DM) process

The research found CRISP-DM is the most commonly used framework for implementing machine learning projects specifically, and it applies to analytics projects as well. Even though CRIPS-DM did not mention anything on gathering, managing and developing the business requirement per say, yet it is focused on the business goals and data gather based on business goal. This sub-process has more explanation on what data to gather or what the processes are involve in which the processes are iterative and could loop back to previous step. The second methodology is TDSP. TDSP methodology is an agile, iterative approach to data science that enables the efficient delivery of predictive analytics solutions and intelligent applications. As shown in Figure 2.7, TDSP provides a lifecycle for data science initiatives to structure project development. TDSP contributes to the enhancement of team communication and learning. This method combines the most appropriate frameworks for development and the processes involve as shown in Table 2.

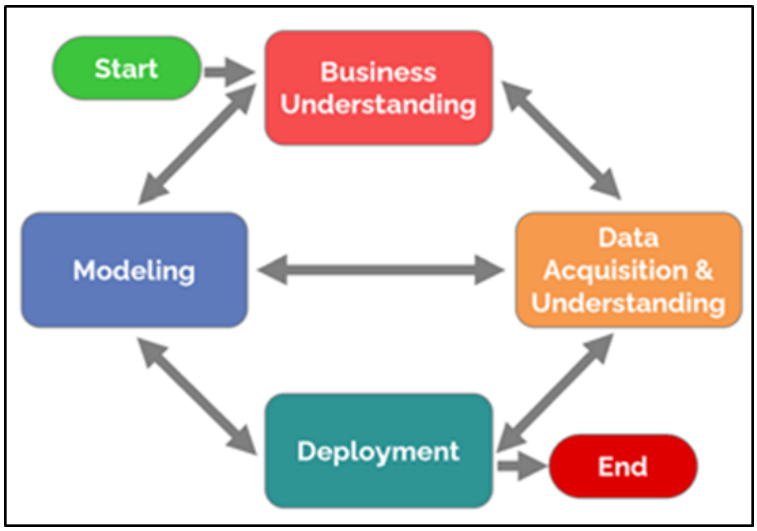

Figure 4.2. TDSP Process

Table 3 TDSP Step and Process

No	TDSP Process	Details
i.	Business understanding	• Business idea. • Defined business perspective. • Identified and evaluated possible scenarios • Project plan.
ii.	Data acquisition and understanding	• Explore data. • Data collection. • Data cleaning. • Data summarization and visualization
iii.	Modelling	• Feature engineering attributes and data features. • Model training machine learning algorithm.
iv.	Deployment	• Produced predictive model and data pipeline. • Real-time or a batch analysis model. • The final data product accredited by the customer.
v.	Deployment	A model is not particularly useful unless the customer can access its results. • Plan deployment. • Plan monitoring and maintenance. • Produce final report. • Review project

This methodology can help support any type of exploratory data science project or any type of improvised analytics effort. However, some of the steps outlined in this paragraph may not be required. TDSP is a team-oriented solution that emphasizes teamwork and collaboration throughout. It recognizes the importance of working as part of a team to deliver Data Science projects. Focus straight to data solution or data prediction model. However, the process to gather data to achieve the data solution and data gathering not clearly explain.

4.3. Requirements Process Related to Big Data Analytics Project in the Malaysia Public Sectors

In Malaysia Public Sector, MAMPU is as a central agency that for the modernization and transformation Public Service Administration has develop process model for the Information Communication and Technology (ICT) project management and application development. Since BDA is part of ICT, it should be covered within the MAMPU's methodology as a guidance for all Government Agency. There are three related methodology related to ICT which are *Pengurusan Projek ICT Sektor Awam* (PPrISA), *Buku Kejuruteraan Sistem Aplikasi Sektor Awam* (KRISA) and *Data Raya Sektor Awam* (DRSA). PPrISA is a methodology for ICT project implementation. The implementation involves 5 phases: Initiation, Planning, Execution and Control and Termination as shown in Figure 3. The PPrISA explain how to implement the ICT project from the starting of the project until project closing. PPrISA did not cater any requirement analysis for the project development [17].

Figure 3. PPrISA Procees [17]

Meanwhile, KRISA ecosystem includes all elements in the ICT environment and its relationship with application system development methodologies and approaches in Government Agencies. KRISA highlight the thing that need a consideration by Government Agency, human resource estimation, change management, stakeholder involvement, software quality assurance, application security and governance [18].

KRISA methodology consists of 6 main phases that need to be understand, use and implement for Government Agency. The phase and explanation are as in Figure 4.

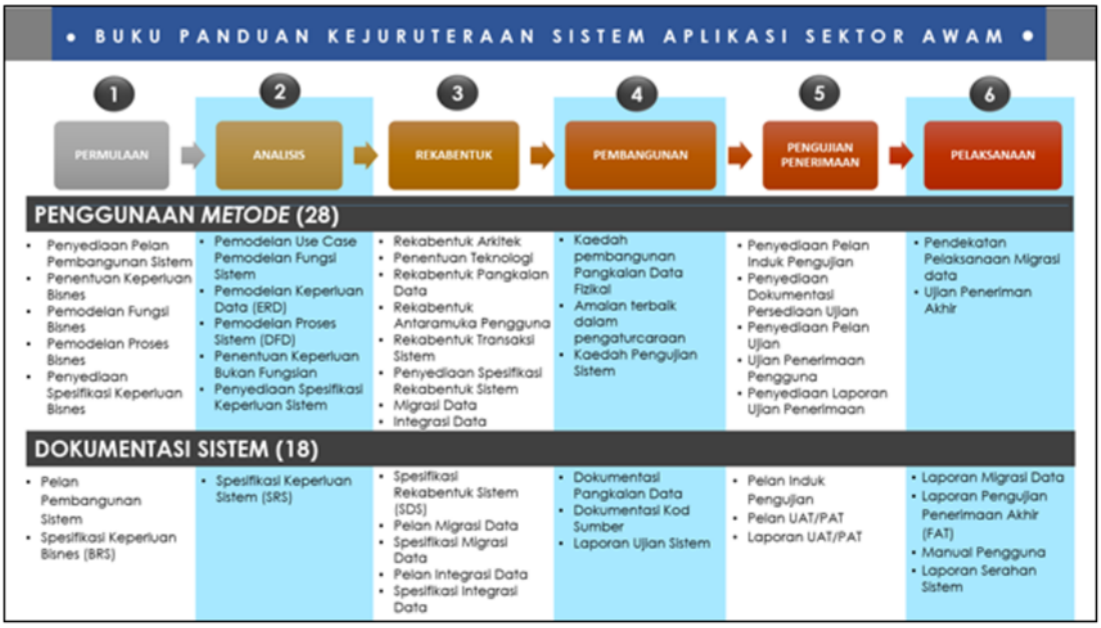

Figure 4. KRISA framework [17]

In the phase 1 of KRISA framework mention about the need of BRS. BRS is a foundation for the development and implementation of the project. This is to understand the organization need and project scope. The development of BRS based on business model in Figure 5.

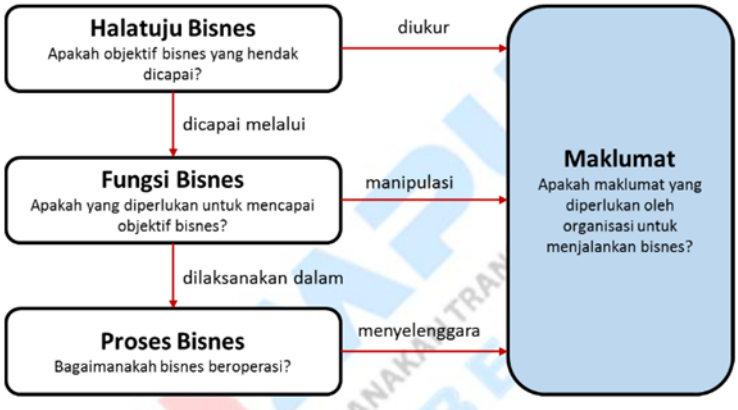

Figure 5. BRS business model in KRISA [18]

BRS start with the understanding of business objective, the business function and business process. KRISA BRS also explain the involvement of relevant stakeholders and step to develop BRS.

Table 4. BRS step in KRISA [18]

Requirement Step	Description
1. Formulate Business Case	• Identify the need for application development
2. Business Understanding	• Project objective • Business process and function • User need • Business change management
3. Choose Subject Matter Expert	• Knowledgeable person on business process • Transfer the knowledge to system application
4. Information Gathering	• Brainstorming • Focus Group • Observation • Interview • Questioner • Joint Application Design (JAD)
5. Business Architecture	• Business documentation that explains about business process in strategic way
6. Information Architecture	• Information environment design • System context, business context and user context
7. Requirement Documentation	• Information gathered are documented

KRISA list the step needed to develop the requirements and not so much about what and how to develop it. To do that, KRISA gather information about the business process, function and user need to develop application. The BRS is for the application and software development that meet user requirement. Eventhough, this process did not particularly cater anything specific to data analytic, still it can be considered to be used within analytics domain. Interestingly, MAMPU has develop the methodology focus on Big Data and Analytics which is DRSA as shown in Figure 6 [19]. It is a methodology and process framework for development and implementation of a Big Data that consists of seven steps which are preparation and planning, stages and high-level implementation, setting up a big data team, documents and deliverables, managing development progress, transitioning to production and post-development monitoring of the project.

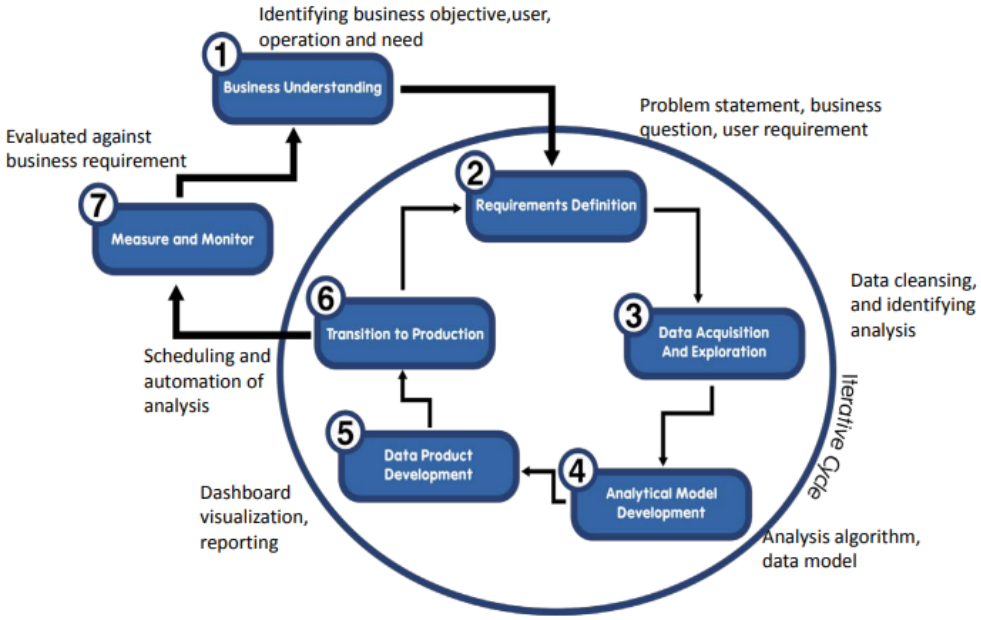

Figure 6. DRSA Big Data Methodology [19]

DRSA methodology is quite similar to ASUM-DM and improvised the iterative cycle which are not being covered in CRISP-DM. In term of analytics requirements, DRSA has second process focus on requirements definition. Within this process, there are a simple template and data processes and workflow. Within the template, there is an example of how a goal and business questions should be written within the analytics project documentation. Meanwhile, data processes and flow show an example of mapping between data sources, analytics processes, business questions and how the visualization should be looked like as the analytics output [17]. However, the sub-processes guidelines, what task to be taken, whose roles and what tools to develop the analytics requirements are still vague in the current methodology.

5.Conclusion

Big Data has risen to the top of the list of company priorities, as it has become a critical foundation for productivity, innovation, and competition in today's market [20]. Poor communication and a lack of clear business goals and objectives have been cited as two of the most common causes of failure in Big Data project. An analytic requirement sets the stage for the BDA project's vision, scope, cost, and schedule. It also lays the groundwork for the project's quality and performance when it is done. To better understand what the business needs are, and to help break down into specific, detailed requirements that can be met. Clear analytics requirements will lead the project well and keep everyone on the same page. If business requirements are not defined clearly, changes, project delays, and extra time and effort are all likely to happen. The Matrix Table 5 will summarise the LR discoveries that have been done at the very beginning of this study. A comparison of the processes, characteristics, and elements used by each methodology will be included.

Table 5. Matrix Summarise the LR Discoveries

Methodology perspectives	RE	CRISP-DM/ ASUM-DM.	TDSP	PPrISA	KRISA	DRSA
Big Data Analytics-Related	Design and management of requirements in software development	Designing data mining projects.	Data science initiatives to structure project development	ICT project implementation	Develop application software	Development and implementation of a Big Data projects
Analytics Requirement	Yes, deep explanation about software requirement (BRS) and not specific to analytics.	Yes, there is black box between business objective and analytics requirements (insights)	Yes, suitable for scientific data analytic project	No. Not specific for analytics requirement	Yes, touched about requirement (BRS) and not specific to analytics.	Yes, it is similar to ASUM)-DM and has template and example for analytics requirements . However, lack of guidelines for requirements process.
Requirement Proccess	1. Identifying System and Context Boundaries. 2. Eliciting Requirements. 3. Documenting Requirements. 4. Requirements Management	1. Determine business objectives 2. Assess situation. 3. Determine data mining goals - define what success looks like from a technical data mining perspective. 4. Produce project plan.	1. Business idea which needs to solved with a machine learning solution. 2. Define Business goals 3. Identified and evaluated possible scenarios. 4. Project plan.	No specific process related to requirement	1. Formulate Business Case 2. Business Understanding 3. Choose Subject Matter Expert 4. Information Gathering 5. Business Architecture 6. Information Architecture: (environment design). 7. Requirement Documentation	No specific process related to requirement

As BDA project BRS are used to analyse and in decision making on health care, economics, education, telecommunications, and many other disciplines. Insightful insights are at the core of BDA sector; hence a clear analytics requirement is used to search for these gems and make sure they are put to good use by organization. The initial step in elicitating, understanding and defining analytics requirements is critical to its success. Some existing models were not suitable to meet the needs of big data analytics condition. Hence a clearly defined analytics requirement process can help in the coordination and organisation of work between the stakeholder, the user, the data scientist, and the analytics developer within the BDA Project.

6. References

[1] Chai, W., Labbe, M., & Stedman, C. (2021, December 14). What is Big Data Analytics and why is it important? SearchBusinessAnalytics. Retrieved July 14, 2022, from https://www.techtarget.com/searchbusinessanalytics/definition/big-data-analytics Marinus J. C. Eijkemans, Houdenhoven, M. v., Nguyen, T., Boersma, E., Steyerberg, E. W. and Kazemier, G. (2010). Predicting the unpredictable: A new prediction model for operating room times using individual characteristics and the surgeon's estimate. Anesthesiology, 1 (12), 41–9. https://doi: 10.1097/ALN.0b013e3181c294c2.

[2] International Institute for Analytics. 2013. "Analytics 3.0." International Institute for Analytics. http://iianalytics. com/a3Liwei, Z., Shoucheng, L., Lidong, W., Lin, X., Jinghui, Y & Guochun, T. (2012). A two-stage approach for surgery scheduling. J Comb Optim. 27, 545–55. https://doi.org/10.1007/s10878-012-9535-2

[3] McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. Harvard business review, 9

[4] S. Soares, The IBM Data Governance Unified Process: Driving Business Value with IBM Software and Best Practices, Ketchum ID:Mc Press, 2010.

[5] Ryan, Lindy (2016). The Visual Imperative: Creating a Visual Culture of Data Discovery

[6] Marr, Bernard. October 27, 2016, 6 Clever Ways Small Businesses Can Use Big Data. Available at http://data-informed.com/6-clever-ways-smallbusinesses-can-use-big-data/

[7] Philbin, S. P. (2015). Applying an Integrated Systems Perspective to the Management of Engineering Projects. Frontiers of Engineering Management, 2(1), 19-3

[8] Wiegers, K. E., & Beatty, J. (2015). Software requirements. Redmond, WA: Microsoft Press

[9] Westfall, Linda. (2006). Software Requirements Engineering: What, Why, Who, When, and How.

[10] Pohl, K., & Rupp, C. (2015). Requirements engineering fundamentals: A study guide for the certified professional for requirements engineering exam, foundation level, IREB compliant. San Rafael, CA: Rocky Nook.

[11] Daneva, M.; Damian, D.; Marchetto, A.; Pastor, O. Empirical research methodologies and studies in Requirements Engineering: How far did we come? Journal of Systems and Software, n. 95, p. 1-9, 2014. http://dx.doi.org/10.1016/j.jss.2014.06.035

[12] Wynekoop, J. L., & Russo, N. L. (1995). Systems development methodologies: Unanswered questions. Journal of Information Technology, 10(2), 65–73. https://doi.org/10.1057/jit.1995.9

[13] Timinger, H., Schmidtner, M., & Reiche, F. (2022). A framework for the construction and tailoring of Engineering Development Process Models. IEEE Transactions on Engineering Management, 1–22. https://doi.org/10.1109/tem.2021.3132699

[14] Vanner, C. (2020, December 15). What is process modeling? 6 essential questions answered, from https://www.bizagi.com/en/blog/process-modeling-and-mapping/what-is-process-modeling-6-essential-questions-answered

[15] Nosrati, M. (2017). Exact requirements engineering for developing business process models. 2017 3th International Conference on Web Research (ICWR). https://doi.org/10.1109/icwr.2017.7959319

[16] Kaur, R., & Sengupta, J. (2011). Software Process Models and Analysis on Failure of Software Development Projects. International Journal of Scientific & Engineering Research, Volume 2(Issue 2).

[17] MAMPU (2018). BAB 1 - Pendahuluan. Retrieved, from https://sqa.mampu.gov.my/index.php/ms/garis-panduan/garis-panduan-pengurusan-projek-ict-prisa

[18] MAMPU (2019). Metodologi Kejuruteraan Sistem Aplikasi Sektor Awam. Retrieved, from https://sqa.mampu.gov.my/index.php/ms/1-5-metodologi-kejuruteraan-sistem-aplikasi-sektor-awam

[19] MAMPU (2015). Analitis Data Raya Sektor Awam (DRSA) - MAMPU. Retrieved, from https://www.mampu.gov.my/ms/data-raya-sektor-awam-drsa

[20] Jensen, M. (2013). Challenges of privacy protection in Big Data Analytics. 2013 IEEE International Congress on Big Data. https://doi.org/10.1109/bigdata.congress.2013.39