

# Design and Usability Evaluation of LDSAM: An Aligned ISO/IEC 29110 – Basic Profile – Development Methodology for Big Data Software Systems in Small Business

David Alejandro Montoya-Murillo <sup>1</sup>, Manuel Mora <sup>2</sup>, Gloria Phillips-Wren <sup>3</sup>,  
Sergio Galván-Cruz <sup>1</sup>, Estela Lizbeth Muñoz-Andrade <sup>1</sup>

<sup>1</sup> Department of Electronic Systems, Autonomous University of Aguascalientes, Aguascalientes, Mexico

<sup>2</sup> Department of Information Systems, Autonomous University of Aguascalientes, Aguascalientes, Mexico

<sup>3</sup> Department of Information Systems, Law and Operations, Loyola University Maryland, Baltimore, USA

Corresponding author: David Alejandro Montoya-Murillo ([david.montoya@edu.uaa.mx](mailto:david.montoya@edu.uaa.mx))

## Editorial Record

**First submission received:**  
September 23, 2025

**Revisions received:**  
December 10, 2025  
February 11, 2026  
March 26, 2026

**Accepted for publication:**  
March 28, 2026

**Academic Editor:**  
Stanislava Mildeova  
University of Finance and Administration,  
Czech Republic

This article was accepted for publication  
by the Academic Editor upon evaluation of  
the reviewers' comments.

**How to cite this article:**  
Montoya-Murillo, D. A., Mora, M., Phillips-  
Wren, G., Galván-Cruz, S., & Muñoz-  
Andrade, E. L. (2026). Design and Usability  
Evaluation of LDSAM: An Aligned ISO/IEC  
29110 – Basic Profile – Development  
Methodology for Big Data Software  
Systems in Small Business. *Acta  
Informatica Pragensia*, 15(2), 440–472.  
<https://doi.org/10.18267/j.aip.317>

**Copyright:**  
© 2026 by the author(s). Licensee Prague  
University of Economics and Business,  
Czech Republic. This article is an open  
access article distributed under the terms  
and conditions of the [Creative Commons  
Attribution License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).



## Abstract

**Background:** System Development Life Cycles (SDLC) are proposed through development methodologies (e.g. Rational Unified Process) and international standards (e.g. ISO/IEC 12207) to guide the systematic development of software products to meet expected time, budget, and functional quality. The ISO/IEC 29110-5-1-2: Software engineering guidelines for the generic Basic profile standard provides a disciplined-systematic lightweight SDLC alternative to agile approaches for businesses interested in ISO/IEC certifications rather than agile ones. However, its utilization in the domain of Big Data Analytics Systems has not been investigated.

**Objective:** Big Data Analytics Systems (BDAS) have been proposed using rigorous SDLCs such as CRISP-DM, Team Data Science Process, and Domino Data Science Lifecycle, but their utilization for small business – or small teams called Very Small Entities (VSEs) – is scarcely reported given the requirements demanding human, technological and financial resources not available in small business. Given this problematic situation, new agile SDLCs for BDAS have been proposed such as Data-Driven Scrum, but some small organizations still require the utilization of a more systematic development process, and thus SDLCs based on ISO/IEC standards are expected.

**Methods:** In this research, the design of Light Data Science – Analytics Methodology (LDSAM) and its pilot usability evaluation from a sample of 27 international practitioners are reported. LDSAM is a lightweight development methodology aligned to the ISO/IEC 29110 – Basic Profile – created especially for VSEs.

**Results:** LDSAM was elaborated using Design Science Research Methodology. Initial usability evaluation results of LDSAM are satisfactory, but further empirical research is encouraged to advance to more mature and stable lightweight SDLCs for BDAS.

**Conclusion:** The LDSAM was successfully developed, and its usability was evaluated by a pilot sample of international academics and professionals. Favorable results were obtained on the usability metrics for the proposed LDSAM SDLC for BDAS.

## Index Terms

Big data analytics systems; BDAS; System development life cycle; ISO/IEC 29110; Very small entities; LDSAM; Light data science – analytics methodology; CRISP-DM; Team data science process; Domino data science lifecycle.

## 1 INTRODUCTION

### 1.1 Context and Motivation of the Research

Software development is commonly guided by the System Development Life Cycle (SDLC), which defines structured processes to transform requirements into functional software products (Olorunshola & Ogwueleka, 2021). An SDLC is typically represented through development models composed of phases, activities, roles, and artifacts, aiming to systematically deliver software within time, budget, and quality constraints (Oktaba & Ibargüengoitia, 1998). Established methodologies such as ISO/IEC 12207 and Rational Unified Process (RUP) have demonstrated their effectiveness in large organizations by improving quality, reducing costs, and optimizing development timelines (ISO/IEC, 2017; Gornik, 2003). However, their direct adoption by Very Small Entities (VSEs) remains challenging due to their organizational complexity and resource demands.

**Table 1.** Summary of the conformance evaluation of the SLDCs for BDAS versus the theoretical SDLC for BDAS.

SDLC element	Theoretical Lightweight SDLC Pro Forma for BDAS of the ISO/IEC 29110 – Basic Profile	CRISP-DM	BDPL	TDSP	DDSL
Roles (7)	User roles: 1. Management roles: 1. Technical roles: 5.	Moderate	Moderate	Moderate	High
	<b>Overall Evaluation of Roles:</b>	<b>ROLES: MODERATE</b>	<b>ROLES: MODERATE</b>	<b>ROLES: MODERATE</b>	<b>ROLES: HIGH</b>
Process - Activities (2, 10)	<b>Process.1 Project Management:</b> (with 4 Activities)	High	High	Very Low	High
	<b>Process.2 Software Implementation: Inception Phases:</b> (with 1 Activities)	High	High	Low	High
	<b>Process.2 Software Implementation: Elaboration Phases:</b> (with 2 Activities)	Moderate	High	Very Low	High
	<b>Process.2 Software Implementation: Construction-Deployment Phase:</b> (with 3 Activities)	High	High	Moderate	Moderate
	<b>Overall Evaluation of Phases-Activities:</b>	<b>PHASE-ACTIVITES: HIGH</b>	<b>PHASE-ACTIVITES: HIGH</b>	<b>PHASE-ACTIVITES: LOW</b>	<b>PHASE-ACTIVITES: HIGH</b>
Products - (22)	<b>Process.1 Project Management:</b> (with 11 products and 1 SSP)	High	Very Low	Very Low	Moderate
	<b>Process.2 Software Implementation: Inception Phases:</b> (with 2 products and 1 SSP)	High	Very Low	Low	Moderate
	<b>Process.2 Software Implementation: Elaboration Phases:</b> (with 6 products and 1 SSP)	High	Very Low	Very Low	Moderate
	<b>Process.2 Software Implementation: Construction-Deployment Phase:</b> (with 7 products and 1 SSP)	High	Very Low	Moderate	Moderate
	<b>Overall Evaluation of Products:</b>	<b>PRODUCTS: HIGH</b>	<b>PRODUCTS: VERY LOW</b>	<b>PRODUCTS: LOW</b>	<b>PRODUCTS: MODERATE</b>
<b>Overall Evaluation of the SDLC:</b>	<b>OVERALL SDLC: HIGH</b>	<b>OVERALL SDLC: MODERATE</b>	<b>OVERALL SDLC: LOW</b>	<b>OVERALL SDLC: HIGH</b>	

In parallel, the ongoing digital transformation across multiple sectors has intensified the demand for Big Data Analytics Systems (BDAS) to support data-driven decision-making in areas such as healthcare, finance, manufacturing, logistics, education, and marketing (Ajah & Nweke, 2019; Wang et al., 2022). Despite their strategic importance, BDAS development presents persistent challenges related to project management, engineering practices, and methodological support (Gorton et al., 2016; Hummel et al., 2018; Davoudian & Liu, 2020). As a result, organizations often rely on ad hoc practices or rigid data science methodologies such as CRISP-DM, KDD, and SEMMA (Chapman et al., 2000; Fayyad et al., 1996; Shafique & Qaiser, 2014; SAS, 2025; Martínez et al., 2019), which were originally conceived for large-scale or exploratory data analysis contexts.

While these methodologies have proven effective in certain scenarios, they generally lack the flexibility and structural balance required by VSEs operating in dynamic market environments (Mohd-Selamat et al., 2018; Gorton et al., 2016). Conversely, agile and lightweight SDLCs, such as ISO/IEC 29110, Scrum, and Extreme Programming, have gained relevance due to their adaptability and reduced process overhead (O'Connor & Laporte, 2017; Sutherland, 2010; Beck, 1999; Hoda, 2018). Nevertheless, their direct application to BDAS remains limited, particularly in projects requiring both methodological rigor and rapid iteration (Boehm & Turner, 2003; Gorton et al., 2016).

Given the need to balance agility with structure, recent studies emphasize the absence of an optimal, standard-backed methodology tailored to BDAS development within VSEs (Mariscal et al., 2010; Plotnikova et al., 2020; Martínez et al., 2021; Haakman et al., 2021; Saltz & Krasteva, 2022; Salazar-Salazar et al., 2024; Montoya-Murillo et al., 2025). In particular, Montoya-Murillo et al. (2025) conducted a systematic literature review comparing heavyweight and lightweight SDLCs for BDAS, concluding that although some approaches align partially with ISO/IEC 29110 – Basic Profile, none provide comprehensive and officially documented definitions of roles, activities, and artifacts suitable for practical adoption by VSEs (see Table 1). For example, CRISP-DM does not formally define roles, while Domino Data Science Lifecycle (DDSL) lacks detailed descriptions of its artifacts (Chapman et al., 2000; Domino Data Lab, 2019).

These findings highlight the need for a lightweight development methodology for BDAS that is both flexible and supported by an international standard. To address this gap, this paper proposes LDSAM, a development methodology based on ISO/IEC 29110 – Basic Profile, integrating elements from CRISP-DM, Team Data Science Process (TDSP), and DDSL (Chapman et al., 2000; Microsoft, 2017; Domino Data Lab, 2019). LDSAM adopts a software engineering perspective grounded in explicit roles, activities, and products (Oktaba & Ibarguengoitia, 1998), enabling VSEs to develop BDAS with recognized quality while optimizing resources and minimizing unnecessary complexity.

## 1.2 Research Objective and Research Questions

The purpose of this study is to conceptually design a development methodology specifically for BDAS. This methodology must be lightweight and adaptable to the nature of a Very Small Entity (VSE), ensuring clarity, efficiency, and scalability in its application. Once defined, it will be documented in an Electronic Process Guide (EPG), designed as an accessible and comprehensive resource for its implementation.

The proposed methodology, called LDSAM, along with the EPG, will undergo an evaluation process through a pilot group consisting of software engineering academics and industry professionals. The evaluation will be based on fundamental criteria such as **Usefulness**, **Ease of Use**, perceived **Value**, and **Intention of Usage**.

For this study, the following research questions (RQs) and their associated null hypotheses (H0) are proposed:

**RQ.1** What elements of Lightweight Development and Big Data Science – Analytics Development Methodologies can be used to elaborate a Lightweight Development Methodology for Big Data Science – Analytics Software Systems that can be evaluated theoretically valid from a Panel of Experts?

**H0.1** There are no elements of Lightweight Development and Big Data Science – Analytics Development Methodologies that can be used to elaborate a Lightweight Development Methodology for Big Data Science – Analytics Software Systems that can be evaluated theoretically valid from a Panel of Experts.

**RQ.2** Can the newly elaborated Lightweight Development Methodology for Big Data Science – Analytics Software Systems be documented in an Electronic Process Guide (EPG) and evaluated as lightweight, useful, easy of using and valuable from a pilot group of Big Data Science – Analytics academics and practitioners?

**H0.2.1** The newly elaborated Lightweight Development Methodology for Big Data Science – Analytics Software Systems cannot be documented in an EPG.

**H0.2.2** The newly elaborated Lightweight Development Methodology for Big Data Science – Analytics Software Systems is not considered useful, easy of using, valuable and utilizable from a pilot group of Big Data Science – Analytics academics and practitioners.

### 1.3 Research Methodology

The scientific research process can be approached through various methodological frameworks (Grunstra et al., 1965). In this study, a combination of three research methods is adopted to generate a comprehensive and optimized methodological framework, enabling the formulation of well-founded conclusions while exploring different alternatives for defining, developing, and evaluating the research problem. The integration of these methods aims to provide a multifaceted perspective, facilitating the construction of a methodology aligned with the project's requirements and offering superior performance compared to the isolated application of a single methodological approach.

In this context, three fundamental methods are considered: the conceptual research method proposed by Mora et al. (Mora et al., 2008), the Design Science Research Methodology (DSRM) developed by Peffers et al. (Peffers et al., 2007), and the research model based on the three cycles of Design Science Research (Hevner, 2007). The final evaluation of the proposed methodological framework will be conducted using the Design Research Evaluation Methods, as detailed in Table 2.

Conceptual research serves as a key source for generating new theories, models, and conceptual frameworks, which, to complete the scientific knowledge cycle, require empirical or deductive validation through other methodological approaches (Blalock, 1969). This type of research is particularly relevant in contexts where study objects cannot be evaluated through physical laws, mathematical models, or established quantitative methods. In the field of information systems, the conceptual approach has become a fundamental tool for formulating new theoretical and methodological models (Mora et al., 2008).

**Table 2.** Design research evaluation methods.

DSRM Steps	Purpose
Step 1) Design problem identification and motivation.	To state the expected overall research goal that delimits the scope of the research, the research questions that focus on the knowledge gaps of interest, and the motivations to pursue the research design. (For these aims is required to conduct a Review of the State of the Art on the specific problem.).
Step 2) Definition of the design objectives and restrictions for the expected artifact.	To define the specific design objectives (i.e. expected qualities in the designed artifact), design restrictions (i.e. the limitations on time, cost and resources utilized to design the artifact), design approach (i.e. analytics, axiomatic or heuristic), design theoretical sources (i.e. the design materials), and design components (i.e. the specific design building blocks).
Step 3) Design and development of the artifact.	To design and implement the expected artifact guided-controlled by the design objectives and restrictions, and using the agreed design approach, design theoretical sources, and design components.
Step 4) Evaluation of the artifact.	To conduct empirical evaluation of the designed and implemented artifact.
Step 5) Communication of research results.	To generate a structured scientific report (i.e. Thesis, Technical Report, Chapter, Conference Proceeding document, or Journal article) of results and communicate them in academic outlets.

## 2 THEORETICAL BACKGROUND

### 2.1 Big Data Analytics Systems

BDAS represents a set of technologies, methodologies, and tools designed to process and analyze large volumes of complex data, commonly known as Big Data. These systems rely on data collection, storage, processing, and analysis, enabling the extraction of patterns, trends, and correlations that are essential for informed decision-making in

various organizational contexts (Pääkkönen & Pakkala, 2015; Tsai et al., 2015). Unlike traditional data analysis methods, BDAS handle structured, semi-structured, and unstructured data from multiple sources, such as IoT sensors, social networks, and smart devices, by applying advanced techniques like predictive models and statistical algorithms (Pääkkönen & Pakkala, 2015; Robinson & Gillis, 2023).

The concept of Big Data's "V's" emerged as a way to describe its fundamental characteristics. Initially, this framework included only three dimensions (volume, velocity, and variety), but over time, it has expanded to encompass five key V's:

1. **Volume:** Refers to the massive amount of data generated daily, ranging from terabytes to petabytes in sectors such as social networks and e-commerce (Robinson & Gillis, 2023).
2. **Velocity:** Describes the speed at which data is generated, processed, and analyzed. In sectors such as healthcare, where real-time medical devices are used, this characteristic is crucial for immediate decision-making (Robinson & Gillis, 2023; Iranmanesh et al., 2023).
3. **Variety:** Represents the diversity of available data types, ranging from text and video to numerical and geospatial data (Robinson & Gillis, 2023).
4. **Veracity:** Refers to the quality and integrity of data, ensuring its reliability for generating meaningful insights (Robinson & Gillis, 2023).
5. **Value:** The final goal of Big Data is to transform data into useful information that drives effective business strategies (Robinson & Gillis, 2023).

Some authors propose a sixth V, **Variability**, which addresses inconsistencies in data flow or usage within organizations (Robinson & Gillis, 2023).

A key aspect of this research is the relationship between BDAS and Very Small Entities (VSE), defined as organizations with small work teams (between 5 and 25 people). These entities face unique challenges when adopting advanced technologies such as BDAS but also have a competitive advantage due to their organizational flexibility and ability to rapidly adapt to new technologies (Salazar-Salazar et al., 2024; Montoya-Murillo et al., 2025; Maroufkhani et al., 2020). Some of the key benefits of using BDAS in VSE include:

- **Adaptive scalability:** BDAS can be implemented in projects with relatively modest data volumes, making them accessible to small businesses with limited resources (Salazar-Salazar et al., 2024; Montoya-Murillo et al., 2025; Kitchin & Lauriault, 2015; Todman et al. 2023).
- **Operational optimization:** They enhance internal efficiency by digitizing processes and utilizing predictive analytics to anticipate market trends (Maroufkhani et al., 2020).
- **Data-driven decision-making:** VSEs can use BDAS to identify business opportunities or improve existing products and services (Maroufkhani et al., 2020; Kitchin & Lauriault, 2015; Todman et al. 2023).

Contrary to common perception, BDAS does not necessarily require large volumes of data to be effective. In contexts where information is limited but well-structured, these tools can provide valuable insights by combining internal sources with publicly available external data (Maroufkhani et al., 2020; Kitchin & Lauriault, 2015; Todman et al., 2023). This is particularly relevant for emerging sectors or companies operating in niche markets.

BDAS are not only essential tools for large corporations but also have a significant impact on VSEs. These entities can implement scalable and customized solutions that maximize their limited resources. Also, the agile approach adopted by many VSEs facilitates the rapid and efficient integration of these technologies into their daily operations (Salazar-Salazar et al., 2024; Montoya-Murillo et al., 2025; Maroufkhani et al., 2020; Todman et al., 2023).

## 2.2 Related Work on BDAS Development Methodologies

LDSAM explicitly addresses the methodological gap identified in the literature. It builds upon the findings of Montoya et al. (Montoya-Murillo et al., 2025), summarized in Table 1, where four SDLCs for BDAS were compared against the ISO/IEC 29110 – Basic Profile standard. The study concluded that The Big Data Project Lifecycle (BDPL) (Lin & Huang, 2018) should be excluded from consideration. This exclusion is justified by the conclusions of Montoya et al. (Montoya-Murillo et al., 2025), which state that the limited availability of documentation related to

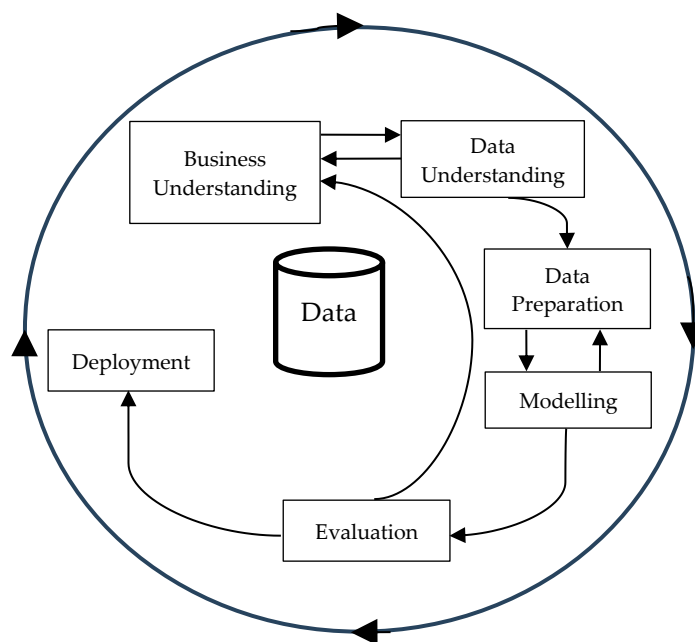
BDPL, together with its dependency on ISO/IEC 15288 (ISO/IEC, 2023), make it incompatible with the lightweight framework required for this research.

Consequently, the subsequent section presents a detailed analysis of the SDLC models that contribute to LDSAM. Each model is described and examined from an SDLC perspective in order to identify the components that are relevant to the proposed methodology.

The analysis conducted in this section does not assume the direct adoption of the reviewed methodologies. Instead, it focuses on identifying and selecting specific methodological elements, such as lifecycle phases, roles, activities, and work products, that are compatible with the constraints of VSEs and can be aligned with the ISO/IEC 29110 – Basic Profile. The selection criteria emphasize minimal process overhead, clarity of responsibilities, support for iterative development, and suitability for production-oriented BDAS. Elements that lack formal definition, introduce excessive complexity, or are weakly supported by documentation are intentionally excluded.

**CRISP-DM** (Chapman et al., 2000; Martínez-Plumed, 2019), an acronym for Cross-Industry Standard Process for Data Mining, is a widely used methodology for developing data analysis and data mining projects. It was created in the 1990s and has become an industry standard due to its structured and flexible approach. Currently, CRISP-DM is the most commonly used methodology in data analysis and data science projects, as it provides an adaptable structure that has served as the foundation for the development of new methodologies, such as Domino Data Lab, Microsoft's TDSP (Martínez-Plumed, 2019, Martínez et al., 2021), and hybrid approaches that integrate Agile and DevOps with data analytics. CRISP-DM does not explicitly define roles. Regarding work products, its phases, tasks, and artifacts can be observed in Table 3. Also, this methodology has been applied in real-world projects (Chapman et al., 2000; Martínez-Plumed, 2019; Martínez et al., 2021). Figure 1 illustrates the CRISP-DM SDLC for BDAS.

From an analytical perspective, CRISP-DM is not adopted as a complete SDLC for LDSAM. Its main contribution lies in the analytical phases related to data understanding, data preparation, modeling, and evaluation, which are well defined and widely validated in practice. These phases are selected due to their strong alignment with the analytical core of BDAS. However, CRISP-DM is not considered suitable as a structural backbone because it does not explicitly define project management activities, formal roles, or standardized software engineering artifacts. Consequently, elements related to governance, lifecycle control, and role definition are intentionally excluded from CRISP-DM in the construction of LDSAM.



**Figure. 1** The CRISP-DM SDLC for BDAS. Based on (Chapman et al., 2000).

**Table 3.** Components of CRISP-DM.

CRISP-DM	Specific elements
Phases	{Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment}
Tasks	{Business Understanding [Determine Business Objectives, Assess Situation, Determine Data Mining Goals, Produce Project Plan], Data understanding [Collect Initial Data, Describe Data, Explore Data, Verity Data Quality], Data preparation [Select Data, Clean Data, Construct Data , Integrate Data, Format Data], Modeling [Select Modeling Technique, Generate Test Design, Build Model, Assess Model], Evaluation [Evaluate Results, Review Process, Determine Next Stages], Deployment [Plan Deployment, Plan Monitoring and Maintenance, Produce Final Report, Review Project]}
Outputs	{Business Understanding [Background , Business Objectives , Business Success Criteria, Inventory of Resources , Requirements Assumptions and Constraints , Risks and Contingencies , Terminology, Costs and Benefits, Data Mining Goals , Data Mining Success Criteria, Project Plan, Initial Assessment of Tools and Techniques], Data Understanding [Initial Data Collection Report, Data Description Report, Data Exploration Report, Data Quality Report], Data Preparation [Rationale for Inclusion/ Exclusion, Data Cleaning Report, Derived Attributes , Generated Records, Merged Data, Reformatted Data, Dataset, Dataset Description], Modeling [Modeling Technique , Modeling Assumptions, Test Design, Parameter Settings , Models, Model Descriptions, Model Assessment , Revised Parameter Settings], Evaluation [Assessment of Data Mining Results w.r.t. Business Success Criteria , Approved Models, Review of Process, List of Possible Actions , Decision], Deployment [Deployment Plan, Monitoring and Maintenance Plan, Final Report, Final Presentation, Experience Documentation]}

**TDSP** (Microsoft, 2017) by Microsoft aims to provide a standardized SDLC for BDAS. This SDLC is presented as an iterative data science methodology designed to systematically develop BDAS, and it claims to be agile, thus categorized as lightweight compared to the heavier CRISP-DM SDLC (Chapman et al., 2000; Martínez-Plumed, 2019). TDSP (Microsoft, 2017) is based on four main components: a definition of the data science lifecycle, a standardized project structure, recommended infrastructure and resources, and necessary tools and utilities for project execution. TDSP is not recommended for very small BDAS projects with a single team composed of a Data Scientist, Data Engineer, and Application Developer. TDSP (Microsoft, 2017) is promoted by Microsoft, suggesting it has been empirically used in real-world BDAS projects. Table 4 presents the components of Microsoft's TDSP for BDAS. Figure 2 shows the TDSP SDLC for BDAS.

TDSP contributes to LDSAM primarily through its explicit definition of roles and its emphasis on operational artifacts that support the deployment of BDAS in industrial environments. Roles such as data scientist, data engineer, and application developer are selected because they complement the role structure defined in ISO/IEC 29110. Also, TDSP artifacts related to solution architecture, deployment reporting, and system monitoring are incorporated to strengthen the transition from experimentation to production. However, TDSP is not adopted in its entirety due to its dependency on specific infrastructure assumptions and its limited suitability for very small teams. Lifecycle elements that introduce organizational or technological overhead are therefore excluded.

**Table 4.** Components of TDSP.

TDSP	Specific elements
Lifecycle	{Business understanding, Data acquisition and understanding, Modeling, Deployment, Customer acceptance}
Tasks	{Business understanding [Define objectives, Identify data sources], Data acquisition and understanding [Ingest the data, Explore the data, Set up a data pipeline], Modeling [Feature engineering, Model training, Model Evaluation], Deployment [Operationalize a Model], Customer acceptance [System Validation, Project hand-off]}
Artifacts	{Business understanding [Charter document, Data sources, Data dictionaries], Data acquisition and understanding [Data quality report, Solution architecture, Checkpoint decision], Modeling [Model], Deployment [A status dashboard that displays the system health and key metrics, A final modeling report with deployment details, A final solution architecture document], Customer acceptance [Exit report of the project for the customer]}
Roles	{Solution architect, Project manager, Data engineer, Data scientist, Application developer, Project lead}

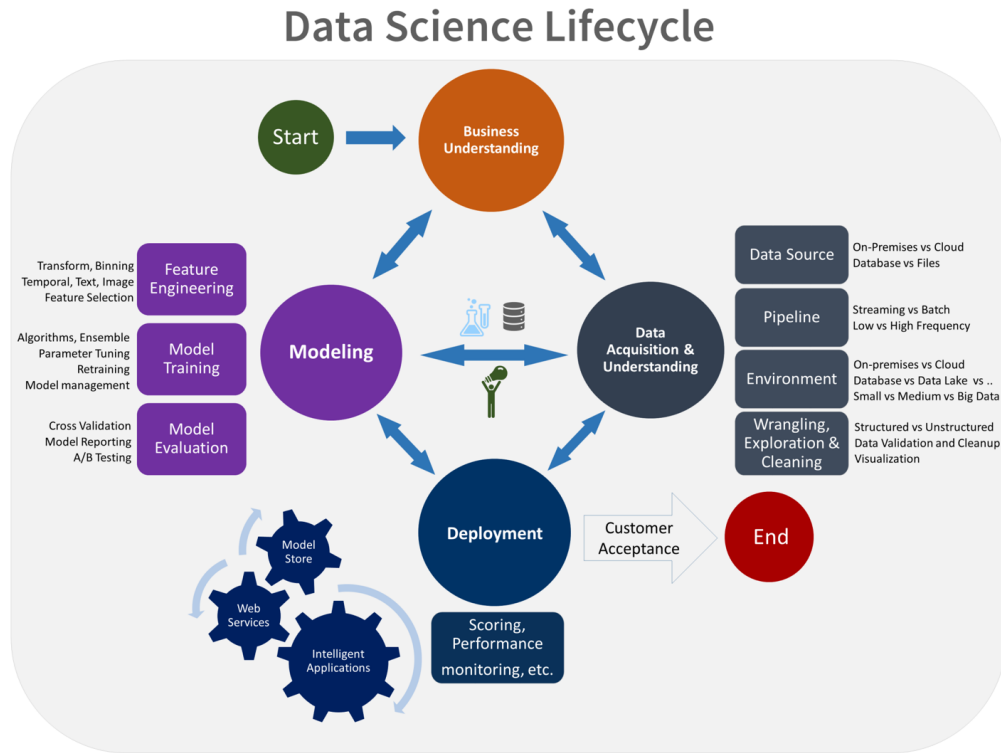


Figure. 2 The Microsoft TDSP SDLC for BDAS. Source: (Microsoft, 2017).

DDSL (Domino Data Lab, 2019) is a full SDLC for BDAS, designed to provide a modern and lightweight SDLC compared to the previous and heavier CRISP-DM SDLC (Chapman et al., 2000), developed by Domino Data Lab. This cycle integrates agile principles and focuses on solving business problems from the start, encouraging collaboration and component reuse. It is structured in six iterative phases that range from ideation to delivery and operations, ensuring a systematic and auditable approach. Similarly to Microsoft TDSP (Microsoft, 2017), no academic reference was found on its utilization, but Domino DSL (Domino Data Lab, 2019) is promoted by Domino Inc., and thus we can infer it has been empirically used in BDAS projects in real-life settings. Figure 3 shows the DDSL SDLC for BDAS. Table 5 presents the components of DDSL by Domino for BDAS.

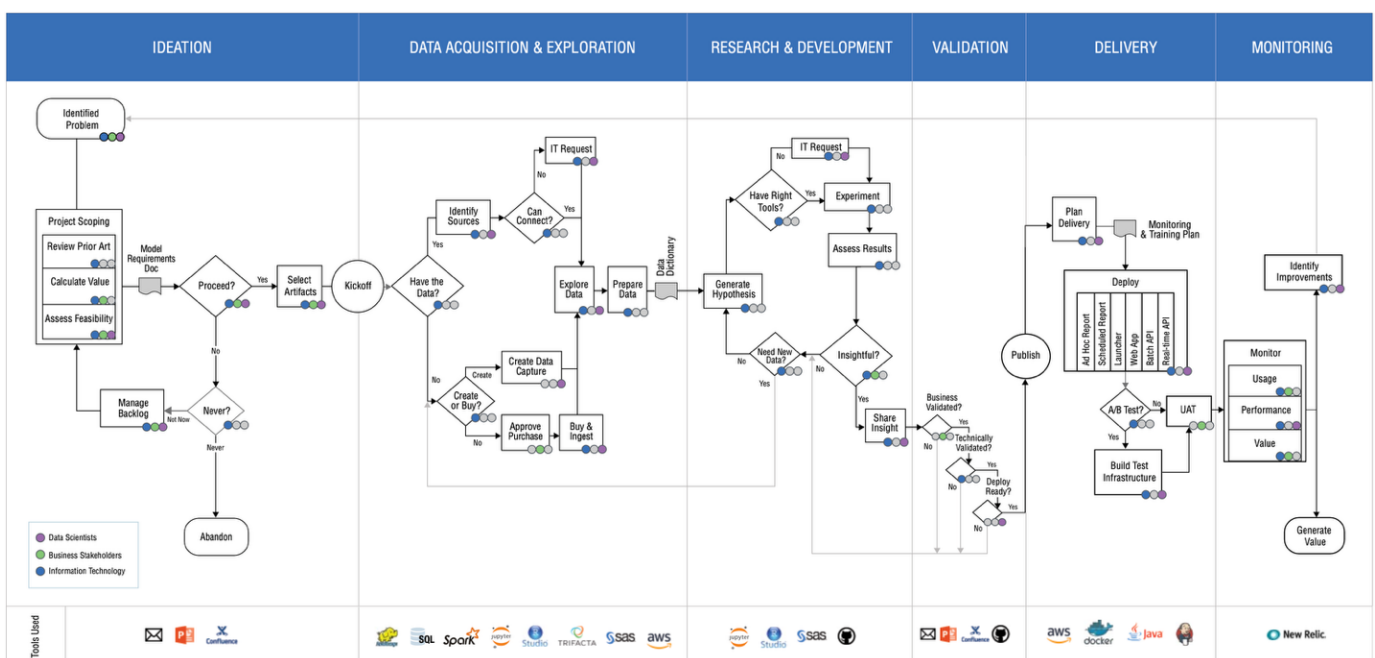


Figure 3. The DDSL SDLC for BDAS. Source: (Domino Data Lab, 2019).

DDSL is analyzed as a complementary source of agile and business-oriented practices rather than as a prescriptive SDLC. Its main contribution to LDSAM is the explicit integration of business value assessment, iterative experimentation, and continuous monitoring throughout the lifecycle. These elements are selected because they reinforce alignment between analytical outcomes and organizational objectives. Nevertheless, DDSL lacks formal standardization and detailed documentation of artifacts, which limits its suitability as a primary methodological framework. For this reason, DDSL elements are selectively incorporated to enhance flexibility and value orientation, while structural and control aspects are delegated to ISO/IEC 29110.

**Table 5.** Components of DDSL.

DDSL	Specific elements
Lifecycle	{Ideation, Data Acquisition and Exploration, Research and Development, Validation, Delivery, Monitoring}
Process	{Ideation [Identified Problem, Project Scoping, Review Prior Art, Calculate Value, Assess Feasibility, Manage Backlog, Select Artifacts], Data Acquisition and Exploration [Getting the Data, Identify Sources the Data, Connect, Create Data (Capture), Buy & Ingest DATA, Explore Data, Prepare Data], Research and Development [Generate Hypothesis, Validate right tools, IT request, Experiment, assess result, Validate the need new Data, Insightful?, Share insight], Validation [Validate the Business, Validate Technically, Validate ready to Deploy, Publish], Delivery [Plan Delivery, Deploy, Test], Monitoring [Monitor, Usage, Performance, Value, Identify Improvements, Generate Value]}
Products	{Ideation [Project Scope document, Project Kick-off, Model Requirements Doc], Data Acquisition and Exploration [Data Dictionary], Research and Development [*Data Model Experiment], Validation [*Validated Data Model], Delivery [*Production Data Model], Monitoring [Monitoring & Training Plan]}
Roles	{Business Stakeholder, Data Product Manager, Data Scientist, Data Infrastructure Engineer, Data Storyteller}

### 2.3 The ISO/IEC 29110 Standard – Basic Profile

The field of Software Engineering has developed a series of standards and process models aimed at facilitating the management and development of high-quality software in organizational environments. The comparative study on Software Process Improvement (SPI) success factors concludes that for successful implementation of models like ISO 9001 or CMMI, "soft" organizational factors are consistently more critical than technical ones. The single most important factor is active commitment and support from senior management, which is essential to drive the initiative. This is reinforced by staff training, effective change management, and clear communication of benefits to ensure real adoption. Therefore, the research emphasizes that ISO 9001 implementation success depends more on leadership and people-centric practices than on the procedural details of the standard itself (Niazi, 2015). The proper implementation of these standards has proven to generate substantial benefits at various levels: cost optimization, improvement in the quality of the final product, reduction in deviations from delivery deadlines, increased user satisfaction, minimization of resource wastage, and greater stability in development processes (Ebert, 2007; Pai et al., 2015). Consequently, the adoption of these models is a widely established practice in large organizations (Ebert, 2007; Pai et al., 2015).

However, these standards have been primarily designed for large-scale corporate environments, limiting their applicability in small businesses due to technical, organizational, and economic constraints. Small businesses typically operate with reduced budgets, less experience in project management, dynamic and informal organizational structures, as well as constant pressure for rapid product delivery (Clarke & O'Connor, 2013; Laporte et al., 2013; O'Connor & Coleman, 2009; Staples et al., 2007). As a result, the adoption of traditional models is hindered by their complexity, the excessive documentation required, the bureaucracy of processes, and the extended development timelines, making them impractical in this context (Clarke & O'Connor, 2013; Laporte et al., 2013; O'Connor & Coleman, 2009; Staples et al. 2007).

To address this issue, the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) have developed the ISO/IEC 29110 series (O'Connor & Laporte, 2017; ISO/IEC, 2011), specifically designed for Very Small Entities (VSE), defined as those with development teams of between five and twenty-five people. This standard provides a balanced approach between agile and traditional methodologies, aiming to improve product quality and process efficiency (O'Connor & Laporte, 2017; ISO/IEC, 2011).

**Table 6.** Quantity of processes (i.e. phases), tasks (i.e. activities), work products, and roles for the four ISO/IEC 29110 standard profiles.

	Entry	Basic	Intermediate	Advanced
Number of processes	2	2	3 (+1 conditional)	3 (+3 conditional)
Number of Tasks	40	67	107 (+ 8 conditional)	120 (+ 24 conditional)
Number of Work Products	14	22	39 (+ 3 conditional)	41 (+ 5 conditional)
Number of Roles	3	7	8 (+ 1 conditional)	8 (+ 1 conditional)

The ISO/IEC 29110 series is structured into different profile groups to adapt to the needs and capabilities of each organization (Laporte et al., 2013). Among these, the Generic Profile Group applies to VSEs that do not develop critical software, and it consists of four levels: Initial, focused on individual projects lasting up to six months; Basic, aimed at individual projects within VSEs; Intermediate, designed for projects with multiple work teams; and Advanced, oriented toward VSEs aiming to become competitive organizations (Laporte et al., 2013). Table 6 summarizes the differences between these profiles in terms of processes, tasks, products, and roles involved.

The flexibility of the ISO/IEC 29110 standard allows its application to various software development life cycles, including waterfall, iterative, incremental, evolutionary, and agile approaches (O'Connor & Laporte, 2017; ISO/IEC, 2011). In Figure 4, adapted from (ISO/IEC, 2011), the phases and activities of the Basic Profile are presented in the context of software engineering.

In the context of this research, the Basic Profile is selected as the most suitable for BDAS projects in small enterprises due to its balance between structure and simplicity. The Initial Profile is discarded, as it does not offer a clear path to organizational certification, and the Intermediate and Advanced Profiles are excluded due to the added complexity of more than 100 specific tasks related to software management and implementation (O'Connor & Laporte, 2017; ISO/IEC, 2011; Laporte et al. 2013). Finally, Table 7 presents the components of the ISO/IEC 29110 - Basic Profile - standard.

**Table 7.** Components of ISO/IEC 29110 – Basic Profile.

ISO/IEC 29110 – Basic profile	Specific elements
Process	{Project Management process, Software Implementation process}
Activities	{Project Management process [Project Planning, Project Plan Execution, Project Assessment and Control, Project Closure], Software Implementation process [Software Implementation Initiation, Software Requirements Analysis, Software Architectural and Detailed Design, Software Construction, Software Integration and Tests, Product Delivery]}
Products	{PM input products [Statement of Work Software Configuration, Software Implementation, Change Request ], PM output products [Project Plan, Acceptance Record, Project Repository, Meeting Record, Software Configuration ], PM internal products [Change Request, Correction Register, Meeting Record, Verification Results, Progress Status Record, Project Repository Backup ], SI input products [Project Plan, Project Repository], SI output products [Software Configuration (Requirements Specification, Software Design, Traceability Record, Software Components, Software, Test Cases and Test Procedures, Test Report, Product Operation Guide, Software User Documentation, Maintenance Documentation ), Change Request ], SI internal products [ Validation Results, Verification Results ]}
Roles	{Analyst, Customer, Designer, Programmer, Project Manager, Technical Leader, Work Team}

Within LDSAM, ISO/IEC 29110 – Basic Profile is selected as the structural backbone of the methodology due to its standardized definition of processes, roles, and work products specifically tailored to Very Small Entities. While CRISP-DM, TDSP, and DDSL contribute domain-specific and operational practices for BDAS development, they do

not provide the level of formalization required for process control, traceability, and reproducibility. ISO/IEC 29110 ensures methodological rigor and compliance, enabling the systematic integration of selected elements from the other methodologies without compromising the lightweight nature of LDSAM.

This section concludes by emphasizing the relevance of the theoretical foundations presented in Tables 3, 4, and 5. From the perspective that an SDLC can be understood as a software development process model (Okta & Iburgüengoitia, 1998), these models provide the necessary adaptability for BDAS projects to align with the components of ISO/IEC 29110 – Basic Profile (Table 7). This alignment will support the clear and specific definition of the phases, activities, functions and products within the LDSAM.

Table 8 presents an analytical comparison of the methodological elements extracted from CRISP-DM, TDSP, DDSL, and ISO/IEC 29110 that were selected for the construction of LDSAM. Table 8 explicitly outlines the selected and excluded elements, together with the rationale behind each decision, emphasizing their suitability for Big Data Analytics Systems developed within Very Small Entities and their alignment with both methodological rigor and practical feasibility.

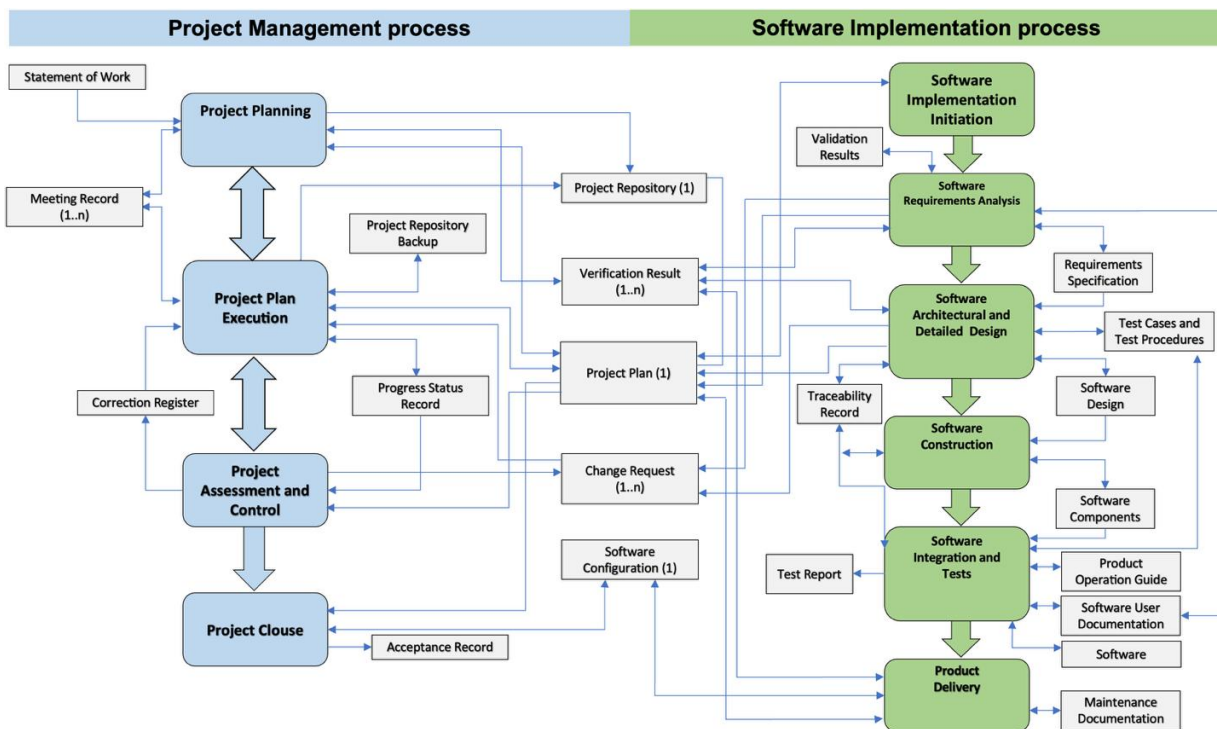


Figure 4. Correlation in the two processes of the ISO/IE 29110 – Basic profile.

Table 8. Analytical selection of methodological elements for LDSAM.

Methodology	Selected Elements	Excluded Elements	Selection Rationale
CRISP-DM	Data Understanding, Data Preparation, Modeling, Evaluation phases; analytical tasks and outputs	Project management activities; formal role definitions; software engineering governance	Strong analytical structure validated in practice; insufficient definition of roles and SDLC control for VSEs
TDSP	Explicit roles (Data Scientist, Data Engineer, Application Developer); deployment and monitoring artifacts; iterative lifecycle	Infrastructure-dependent practices; organizational structures unsuitable for very small teams	Provides operational rigor and role clarity; partial adoption avoids excessive overhead
DDSL	Business value assessment; iterative experimentation; continuous monitoring practices	Lack of standardized artifacts; absence of formal lifecycle control	Enhances agility and value orientation; not suitable as a standalone standardized SDLC
ISO/IEC 29110 – Basic Profile	Project Management and Software Implementation processes; formal roles; standardized work products	Intermediate and Advanced profiles; excessive task granularity	Standard-backed structure tailored for VSEs; provides methodological backbone with controlled complexity

## 3 LDSAM LIGHT DATA SCIENCE – ANALYTICS METHODOLOGY – DESIGN, DESCRIPTION AND CONCEPTUAL VALIDATION

### 3.1 Definition of the design objectives

The expected Design Objectives (DOs) are:

1. The designed artifact provides a lightweight (i.e. responsive, flexible, speedy, lean, simple, lightweight, and fine-grain documented (Conboy, 2009; Qumer & Henderson-Sellers, 2008) workflow (i.e. a value stream) for designing, building, and implementing a new minimum viable Light Data Science - Analytics Methodology.
2. The designed artifact is useful, ease to use, and valuable (Galvan et al., 2021) for small companies, software developers, and IT practitioners.
3. The designed artifact is fine-grain documented including the roles-set component, phases-activities set component, and artifacts-templates-set component.

### 3.2 Design Restrictions

Design Restrictions (DRs) require consideration of parameters such as time, budget, theoretical sources, and available software. The DRs that were agreed upon are:

1. The designed artifact must be composed of design building-blocks from relevant design theoretical sources (DTSs).
2. The designed artifact must be designed in a short-term period (at most 6 months) and under the assigned research budget.
3. The designed artifact must be documented in an Electronic Process Guide.

### 3.3 Design theoretical sources.

The Design Theoretical sources (DTSs) are the key sources of the design components that will be chosen to create the artifact. The DTSs selected were proposed based on the theoretical background, which are mentioned in Table 9.

**Table 9.** *The design theoretical sources.*

DTS.1	ISO/IEC 29110 – Basic profile (ISO/IEC, 2011)
DTS.2	CRISP-DM: Cross Industry Standard Process for Data Mining (Chapman, 2000)
DTS.3	TDSP: The Team Data Science Process (Microsoft, 2017)
DTS.4	DDSL: Domino Data Science Lifecycle (Domino Data Lab, 2019)

Every single element such as Roles, Activities, and Artifacts for the DTS will be considered and discussed with the team to get the Design Components.

### 3.4 Design components for the expected artifact.

Evaluating carefully the DTS, we have selected the potential design components (DCS) to be used in the design of the artifact. Some components may not be used in the final design.

Table 3, Table 4, Table 5 and Table 7 have all the Design Components selected from the four DTSs by the research team based on their experience and expertise. An iterative process was used to identify the most important components to design the artifact.

Note that although the following section presents the heuristic analysis conducted across the three iterations of component selection, a preliminary analysis was carried out in this phase to detect functional patterns across the three BDAS methodologies. This analysis was performed through a systematic comparison of their activities, which enabled the empirical identification of six key activities for a BDAS. The components associated with these activities will be selected in the subsequent step, through a process of inclusion, exclusion, or unification, depending on the requirements of each task, role, and product.

### 3.5 Design and Development of the Artifact

To design the BDAS Methodology the research team applied Means-Ends Analysis heuristic (Newell & Simon, 2019; Greeno et al., 1988) in four steps:

- **Step 1.** To represent the design problem defining an initial state  $S_i$ , a pursued final state  $S_f$ , a set of heuristic operators  $\{HO_x(S_y, S_z), \dots\}$  that can transform the state  $S_y$  to the state  $S_z$ , a set of design objectives  $\{DO_j, \dots\}$  and design restrictions  $\{DR_k, \dots\}$  expected to be satisfied by the final state  $S_f$ , and two qualitative functions  $EvalDOs(DO's)$  and  $EvalDRs(DR's)$  to evaluate the logical satisfaction of  $DO's$  and  $DR's$ .
- **Step 2.** To set up the initial state  $S_i$  and the pursued final state  $S_f$ , and determine the initial qualitative evaluations  $EvalDOs(DO's)$  and  $EvalDRs(DR's)$  for the initial state  $S_i$  and the pursued final state  $S_f$ .
- **Step 3.** To apply a sequence of heuristic operators  $\{HO?(S_i, S_2); HO?(S_2, S_3); \dots; HO?(S_n, S_f)\}$  based on a logical analysis of the operators that can transform the initial state  $S_i$  in the pursued final state  $S_f$ .
- **Step 4.** To evaluate the level of compliance of the pursued final state  $S_f$ , regarding the design objectives  $\{DO_j, \dots\}$  and design restrictions  $\{DR_k, \dots\}$ .

Table 3, Table 4, Table 5 and Table 7 present the Design Components selected for the first iteration. The research team evaluated and discussed the usefulness of each component. Once the Design Components were defined, the second iteration of the process reviewed each component and analyzed its relevance for inclusion in the BDAS Methodology. This second iteration identified a new set of Design Components that could be essential for the methodology.

Across all three iterations, component exclusion decisions were explicitly guided by redundancy with ISO/IEC 29110 elements, misalignment with the defined Design Objectives and Design Restrictions, or the introduction of unnecessary methodological overhead for Very Small Entities. This ensured a progressive convergence toward a minimal, coherent, and standards-aligned methodology.

In the third and final iteration, the minimum necessary Design Components for the Methodology were determined. During this process, components already covered by DTS.1 (ISO/IEC 29110 - Basic Profile-) [32] were eliminated, and additional components from other DTS [12,33,34] were incorporated. Appendix A presents the Design Components selected in each of the three iterations for the BDAS Methodology, including roles (Appendix A.1), activities (Appendix A.2), tasks (Appendix A.3), and products (Appendix A.4).

The design was concluded by fully adopting the ISO/IEC 29110 – Basic Profile – standard to ensure that LDSAM retained the benefits of the standard by aligning with it. Regarding the roles, the exact ones from the standard were used, but specifications were added to enhance their correlation with BDAS. As for the processes, the same ones were maintained, such as Project Management and Software Implementation. Regarding the activities, six new ones were added, which are complementary to the ISO/IEC 29110 – Basic Profile – to provide a greater focus on projects where the key aspect is the data (the added activities can be seen in Figure 5). For the required products in BDAS projects, when using the standard, a terminology was adopted in which we referred to them as internal products, to which 11 new ones were added.

The Lightweight Data Science Analytical Methodology (LDSAM), formalized as the ISO/IEC 29110 extension for BDAS, provides substantial methodological innovations by establishing a verifiable conduit between standardized software engineering practice and the specialized requirements of data-centric projects. Unlike the original ISO/IEC 29110 Basic Profile, which lacks provisions for iterative analysis and model management, the LDSAM formally integrates the six canonical phases of analytical frameworks (e.g., CRISP-DM). This integration is sustained by defining specific Internal Products, such as the Data Cleaning Report and the Model Assessment artifact, which demand the documentation of empirical evidence throughout the data lifecycle. These mandated artifacts ensure that the selection and validation of data assets and predictive models meet rigorous quality and traceability standards, an essential layer of governance critically absent when traditional software engineering models are directly applied to Big Data Analytics Systems.

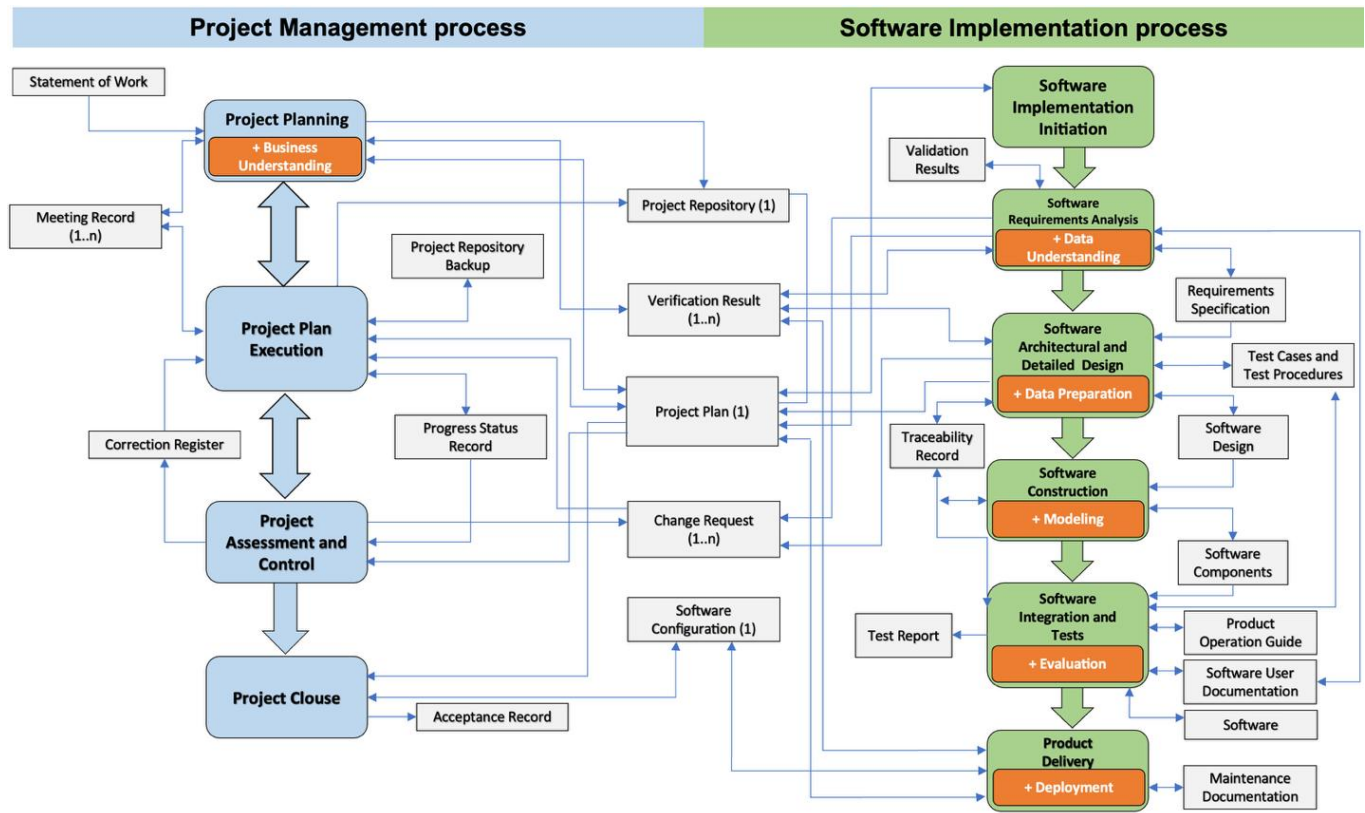


Figure 5. Diagram of ISO/IEC 29110 – Basic Profile – for BDAS +.

The LDSAM, formalized as an extension of ISO/IEC 29110 for BDAS, introduces significant methodological innovations by establishing a verifiable link between standardized software engineering practices and the specialized requirements of data-driven projects. Unlike the original ISO/IEC 29110 Basic Profile, which does not provide explicit provisions for iterative analysis or model management, LDSAM formally incorporates the six canonical phases of analytical frameworks (e.g., CRISP-DM).

This integration is supported by the definition of specific internal products, such as the data cleansing report and the model evaluation artifact, which require the systematic documentation of empirical evidence throughout the data life cycle. In addition, LDSAM seeks to offer a complete specification of roles, processes and activities, and associated products, ensuring the level of methodological traceability necessary for execution. To facilitate operational adoption, the methodology also provides templates for all deliverables, which are accessible for consultation and download directly from the EPG.

The presence of these mandatory artifacts ensures that the selection and validation of data assets and predictive models conform to rigorous standards of quality and traceability. This layer of governance is essential and is critically lacking when traditional software engineering models are applied directly to big data analytics systems.

### 3.6 Design of the Electronic Process Guide for ISO/IEC 29110 – Basic Profile – for BDAS +

Previously, we reported that we had chosen to design the new EPG for LDSAM, which we named ISO/IEC 29110 - Basic Profile - for BDAS + EPG. However, before constructing the EPG artifact, it was necessary to have the source content and an expected structure for the EPGs.

To design a methodology that is considered by software developers as lightweight (neither agile nor rigorous) useful, easy of using and valuable, based on the best practices provided by the ISO/IEC 29110 - Basic Profile, and adding the highlighted BDAS characteristics from other methodologies, the Roles, Processes and Activities, and Products are described in the tables of section 2.2 Related Work on BDAS Development Methodologies.

This process was meticulously applied by the principal investigator and discussed in a team with the second investigator (doctoral advisor) and the external doctoral advisor. Several iterations were necessary to adjust the overall levels. Finally, a document titled ISO/IEC 29110 - Basic Profile for BDAS+ was created. This document was also evaluated, and the results were satisfactory, allowing the development of the ISO/IEC 29110 - Basic Profile for BDAS + EPG.

### 3.7 Elaboration of the Electronic Process Guide for ISO/IEC 29110 – Basic Profile – for BDAS +

The ISO/IEC 29110 – Basic Profile – for BDAS + EPG was developed in Visual Studio Code using HTML, CSS and JavaScript (see basic description in Figure 6). This final artefact ISO/IEC 29110 – Basic Profile – for BDAS+ EPG is available for consulting from the corresponding author.

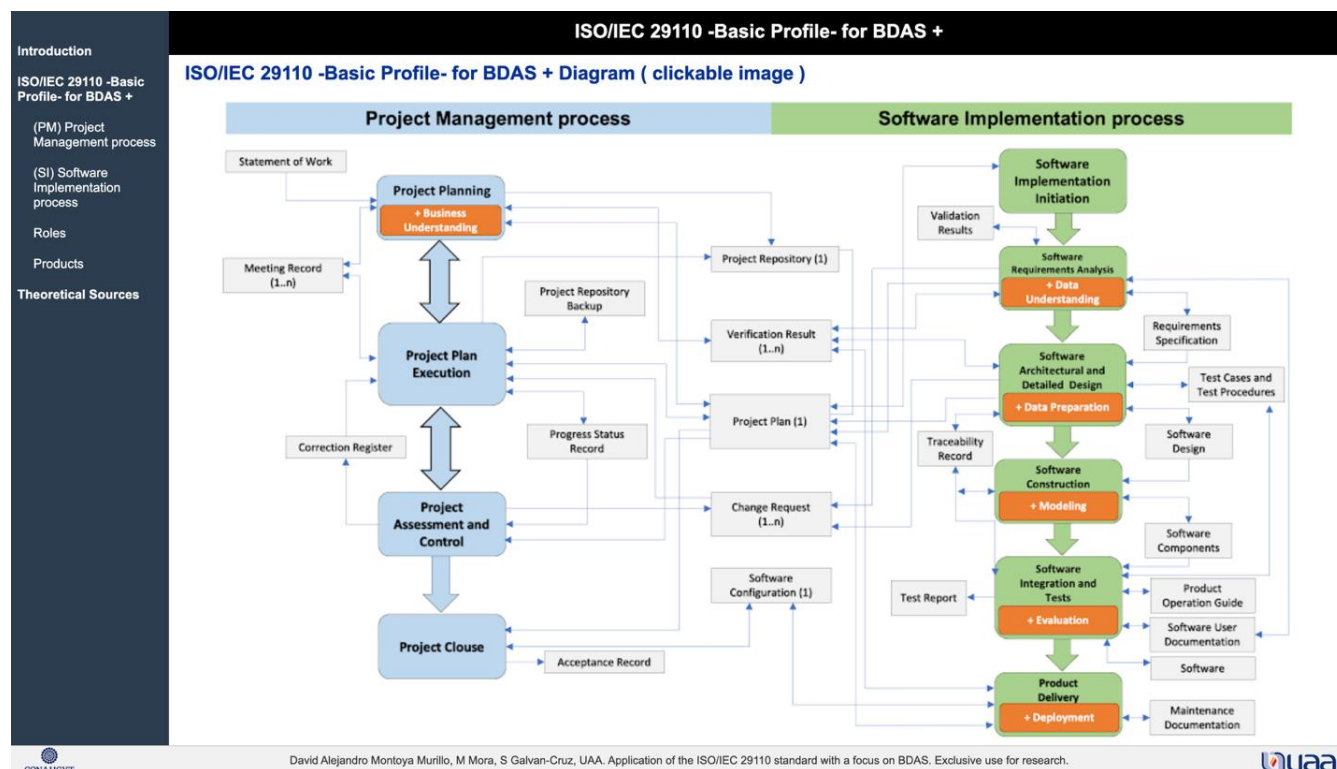


Figure 6. EPG – ISO/IEC 29110 – Basic Profile – for BDAS +.

The presented artifact constitutes a formal methodological extension of the international standard ISO/IEC 29110 – Basic Profile, specifically engineered for the management and implementation of BDAS projects within VSEs (organizations comprising a maximum of 25 participants). A brief explanation of the components of the EPG is presented below:

#### 1. (Introduction): Rationale and Context

The proposed model addresses the high failure rate commonly reported in Data Science projects. The primary objective is to establish a SDLC that harmonizes the discipline and quality assurance inherent in an international standard with the lightness and specialized phases requisite for data analytics. The projected duration for applying this SDLC to BDAS projects is framed within a timeframe of 6 to 36 weeks.

#### 2. (ISO/IEC 29110 - Basic Profile - for BDAS +) Structural Components and Extensions (BDAS+)

The BDAS+ extension preserves the process duality (PM and SI) of the ISO/IEC 29110 standard while enriching it through three primary categories of modifications, includes an integrated diagram illustrating the extended flow of information and the interrelationships between the augmented PM and SI processes:

## 2.1. Core Data Analytics Phases

Six fundamental activities from the BDAS domain are integrated into the workflow, specifically mapped to and complementing the Project Management (PM) process and Software Implementation (SI) process activities:

### 2.1.1 BDAS Phases (Integrated into PM):

- Business Understanding: Clear definition of business objectives and project success criteria.

### 2.1.2 BDAS Phases (Integrated into SI):

- Data Understanding: Exploration and validation of data quality and relevance for analysis.
- Data Preparation: Data cleaning, transformation, and feature selection for model training.
- Modeling: Development, training, and tuning of analytical and predictive models.
- Evaluation: Assessment of model performance against established business and statistical metrics.
- Deployment: Integration of the approved model into the VSE's operational environment.

## 2.2. Roles and Diagrams

The artifact incorporates additional roles reflecting the multidisciplinary needs of a BDAS project (e.g., Data Scientist, Data Engineer).

## 2.3. Artifacts (Internal Products)

To ensure the traceability and quality mandated by the standard, "internal products" of a technical and analytical nature are specified. These artifacts detail the BDAS workflow within the existing ISO/IEC 29110 work products. Examples include the Data Cleaning Report, documentation of the Modeling Technique, and Model Assessment reports. In addition, at the beginning of the section describing the products, a direct access is provided for downloading the corresponding templates ([Link to download products template](#)). This feature facilitates the practical adoption of LDSAM by ensuring that project stakeholders have access to the required standardized formats from the outset of the process.

## 3. (Theoretical Sources) Structure and Theoretical Foundations

The BDAS+ methodological model is founded upon the integration of four key Design Theoretical Sources (DTs) to ensure both theoretical robustness and practical relevance:

1. ISO/IEC 29110 - Basic Profile: Provides the foundational Software Engineering framework, defining the Project Management (PM) and Software Implementation (SI) processes.
2. CRISP-DM (Cross Industry Standard Process for Data Mining): Contributes the canonical phase structure for data mining and analysis projects.
3. TDSP (The Team Data Science Process): Provides collaborative practices and workflow design for data science teams.
4. DDSL (Domino Data Science Lifecycle): Complements the framework with a focus on model governance and scalability.

Consequently, the EPG artifact establishes a verifiable and coherent framework for implementing advanced analytics solutions in resource-constrained environments.

## 3.8 Evaluation of LDSAM

To validate RQ.2, which was evaluated as useful, easy to use and valuable by a pilot group of academics and professionals from the disciplines of Big Data, Data Science-Analytics or experience in the development or use of BDAS, a validation procedure was applied, adapted from the senior evaluator panel methodology described in (Beecham, 2005; Saadatmand, 2024; Hakim et al., 2024). Following the senior evaluator panel approach described by Beecham (2005), Saadatmand (2024), and Hakim et al. (2024), the validation focused on expert-based conceptual assessment rather than empirical field deployment. These methodologies informed the definition of explicit inclusion criteria to identify senior evaluators, the use of structured questionnaires to assess conceptual validity and usability, and the filtering of evaluators prior to analysis. This approach was adapted to the context of Big Data

Analytics Systems and Very Small Entities by employing a lightweight, fully digital validation process and statistical techniques suitable for small samples. This procedure consisted of the following tasks:

**Task 1. Generate the artifact to be validated.**

The target artifact was the ISO/IEC 29110 – Basic Profile for BDAS+ EPG, previously introduced and described in Section 3.7 of this article.

**Task 2. Define the demographic questionnaire and the conceptual validity questionnaire.**

Both instruments were adapted from (Mora et al., 2009). The demographic questionnaire contains 8 general questions and 6 specific questions, aimed at identifying senior evaluators according to the validation metrics. Appendix B contains the demographic questionnaire.

The conceptual validity questionnaire contains 7 items designed to measure theoretical validity (3 items) and theoretical coherence (4 items). A five-point Likert scale was used, ranging from 1 (strongly disagree) to 5 (strongly agree). Appendix C contains the conceptual validity questionnaire.

The usability questionnaire includes four dimensions: Usefulness, Ease of use, Value, and Attitude. A five-point Likert scale was used for the first three dimensions, while Attitude employed a semantic differential scale (-3 to +3). Appendix D contains the usability questionnaire.

**Task 3. Establish inclusion criteria for senior evaluators.**

The qualitative criteria were defined as follows:

- IC.1) Senior evaluators must have a reasonable number of years using and/or teaching practices related to international standards and/or BDAS development practices.
- IC.2) Senior evaluators must have participated in a reasonable number of projects involving international standards and/or BDAS development projects.
- IC.3) Senior evaluators must consider themselves as having high or very high experience in international standards and/or BDAS development practices.

These qualitative criteria were assessed through the final six specific questions of the demographic questionnaire.

**Task 4. Create the digital material package.**

The research team assembled the following materials: validation instructions, the ISO/IEC 29110 – Basic Profile for BDAS+ EPG, the demographic questionnaire (Google Form), the conceptual validity questionnaire (Google Form), and the usability questionnaire (Google Form).

**Task 5. Define and contact the list of potential evaluators.**

The research team contacted 27 international academics and practitioners from the fields of software engineering, international standards, and BDAS.

**Task 6. Filter the set of senior evaluators according to the inclusion criteria.**

From the 27 contacted individuals, 6 met the senior evaluator inclusion criteria.

**Task 7. Report demographic data.**

Table 10 (part 1 and part 2) in presents a summary of the demographic data for the 27 contacted evaluators, grouped into junior and senior evaluators.

**Task 8. Compute and report statistical metrics.**

Present the reliability, convergent validity, and discriminant validity metrics of the conceptual validity questionnaire, computed using the full dataset of 27 evaluators. For the usability calculations, the PLS statistical technique was applied (Vinzi et al., 2010). PLS is suitable when determining the validity and reliability of measurement models with small sample sizes and without requiring normally distributed data (Vinzi et al., 2010; Gefen et al., 2000; Hair, 2019; Russo, 2021).

**Table 10.1** Demographic data for the 27 evaluators blocked by junior and senior evaluators. (Part 1)

Demographic Item	Junior Evals. Block %	Senior Evals. Block %
<b>1. Age range:</b>		
<=30 years	18.5%	0.0%
31-40 years	29.6%	3.7%
41-50 years	14.8%	0.0%
> 50 years	14.8%	18.5%
<b>2. Academic highest gained level:</b>		
Bachelor level	18.5%	0.0%
Bachelor enhanced with Professional Certifications	18.5%	7.4%
Master level	22.2%	3.7%
Doctorate level	18.5%	11.1%
<b>3. Main area of formal studies:</b>		
Computer Engineering	40.7%	3.7%
Business Informatics	18.5%	14.8%
Business Management	3.7%	0.0%
Other	14.8%	3.7%
<b>4. Main work setting:</b>		
Business enterprise	44.4%	11.1%
University/Research Unit	25.9%	11.1%
Government Unit	7.4%	0.0%
<b>5. Scope of work setting:</b>		
Regional	40.7%	0.0%
Nationwide	18.5%	7.4%
Worldwide	11.1%	22.2%
<b>6. Region of working setting:</b>		
USA/CAN	51.9%	0.0%
Europe	22.2%	18.5%
Asia	0.0%	0.0%
Latin America	3.7%	3.7%
<b>7. Years in work settings:</b>		
1-5 years	22.2%	0.0%
6-10 years	18.5%	3.7%
11-15 years	11.1%	0.0%
16-20 years	7.4%	0.0%
20 or more years	18.5%	18.5%
<b>8. Main work position:</b>		
Academic/Researcher	29.6%	11.1%
IT Project Manager / IT Consultant	29.6%	11.1%
Business Manager / Business Consultant	7.4%	0.0%
IT Senior Developer	11.1%	0.0%

**Table 10.2** Demographic data for the 27 evaluators blocked by junior and senior evaluators. (Part 2)

Demographic Item	Junior Evals. Block %	Senior Evals. Block %
<b>9A. Years involved (i.e. knowing, using, teaching, investigating or giving consulting) on LIGHTWEIGHT PROCESS (Disciplined Agile, ISO/IEC 29110 standard, RUP for Small Projects, MSF for Small Projects, or Hybrid Scrum-XP):</b>		
<1 year	40.7%	0.0%
1-3 years	22.2%	0.0%
4-6 years	11.1%	0.0%
7-9 years	0.0%	7.4%
10 or more years	3.7%	14.8%
<b>9B. Years involved (i.e. knowing, using, teaching, investigating or giving consulting) on Data Science Analytics Systems:</b>		
<=5 years	55.6%	7.4%
6-10 years	14.8%	0.0%
11-15 years	3.7%	0.0%
16-20 years	0.0%	3.7%
>20 years	3.7%	11.1%
<b>10A. Number of projects (academic, training or consulting ones) involved with on LIGHTWEIGHT PROCESS (Disciplined Agile, ISO/IEC 29110 standard, RUP for Small Projects, MSF for Small Projects, or Hybrid Scrum-XP):</b>		
1-3	63.0%	0.0%
4-6	11.1%	7.4%
7-9	0.0%	3.7%
10 or more	3.7%	11.1%
<b>10B. Number of projects (academic, training or consulting ones) involved on Data Science Analytics Systems:</b>		
1-3	51.9%	7.4%
4-6	22.2%	0.0%
7-9	0.0%	0.0%
10 or more	3.7%	14.8%
<b>11A. Self-evaluation on the expertise level on LIGHTWEIGHT PROCESS (Disciplined Agile, ISO/IEC 29110 standard, RUP for Small Projects, MSF for Small Projects, or Hybrid Scrum-XP):</b>		
very high level of expertise	3.7%	7.4%
high level of expertise	18.5%	7.4%
moderate level of expertise	18.5%	7.4%
low level of expertise	29.6%	0.0%
very low level of expertise	7.4%	0.0%
<b>11B. Self-evaluation on the expertise level on Data Science Analytics Systems:</b>		
very high level of expertise	3.7%	11.1%
high level of expertise	25.9%	7.4%
moderate level of expertise	18.5%	3.7%
low level of expertise	25.9%	0.0%
very low level of expertise	3.7%	0.0%

### 3.8.1 Conceptual Evaluation of LDSAM

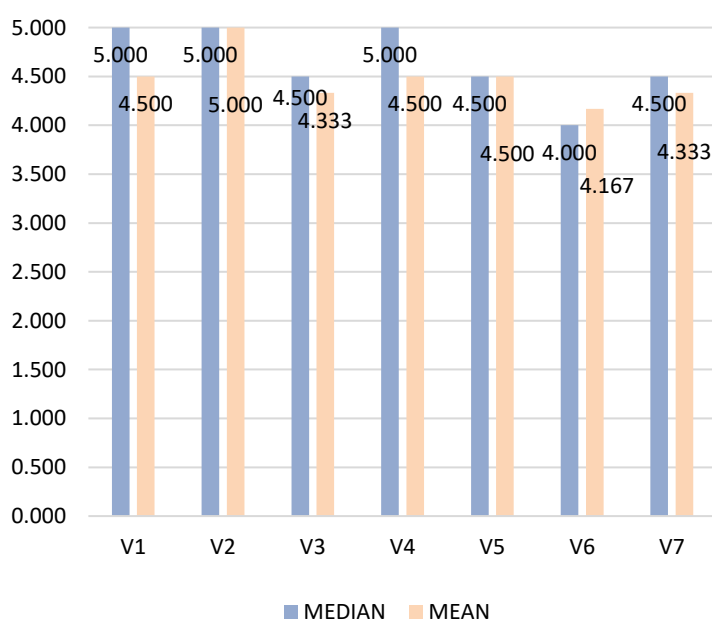
To calculate mean and standard deviation of each item in the questionnaire. The mean and standard deviation are reported in Table 11 It was used a Likert scale from 1 (total disagreement with asked item) to 5 (total agreement with asked item). Where according to Task 3. Establish inclusion criteria for senior evaluators, 6 of the 27 were finally selected as Senior. In Figure 7, the Median and Mean of the Seniors are represented in the 7 items.

**Table 11.** Mean and Standard Deviation of the Constructs/Items C1 and C2.

CONSTRUCT / ITEMS	MEAN	STANDARD DEVIATION
<b>C1 THEORETICAL VALIDITY</b>	4.750	0.622
ITEM#1. The conceptual product is supported by robust theoretical knowledge (e.g. based on scientific literature).	4.500	0.837
ITEM#2. The theoretical knowledge used for elaborating this conceptual product is relevant for the addressed topic.	5.000	0.000
<b>C2 THEORETICAL CONSISTENCY</b>	4.367	0.718
ITEM#3. The scientific literature considered for elaborating this conceptual product does not present important omissions for the topic.	4.333	0.816
ITEM#4. The conceptual product is logically coherent.	4.500	0.837
ITEM#5. The conceptual product is adequate for achieving the purpose of its utilization.	4.500	0.548
ITEM#6. The conceptual product provides new scientific-based knowledge (e.g. it is not a just a duplication of an existent conceptual product).	4.167	0.753
ITEM#7. The conceptual product is supported by robust theoretical knowledge (e.g. based on scientific literature).	4.33	0.816

The results indicate a highly positive evaluation of the conceptual product, with high mean values across all items, ranging between 4.167 and 5.00. The construct C1 Theoretical Validity presents an overall mean of 4.75, reflecting a very favorable perception regarding theoretical support, with standard deviations between 0.000 and 0.837, suggesting low to moderate variability among evaluators. For the construct C2 Theoretical Consistency, the overall mean is 4.367, indicating a good perception of internal coherence. Within this construct, the items addressing logical coherence and adequacy for the intended purpose reach mean scores of 4.50, while the lowest mean corresponds to ITEM #6 (4.167), although it still falls within a positive evaluation range. The standard deviations for this construct vary between 0.548 and 0.837, indicating a reasonable degree of homogeneity in the responses.

Importantly, the predetermined criterion established that the mean score for constructs C1 and C2 should be less than or equal to 3.00 to be declined. Since all mean values clearly exceed this threshold, the conceptual product is accepted in terms of both theoretical validity and theoretical consistency. Overall, the results provide evidence of strong theoretical validity and consistency, with controlled variability and no indications of extreme values.



V1. The conceptual product (ISO/IEC 29110 -Basic Profile- for BDAS +) is supported by robust theoretical knowledge (e.g. based on scientific literature).

V2. The theoretical knowledge used for elaborating this conceptual product (ISO/IEC 29110 -Basic Profile- for BDAS +) is relevant for the addressed topic.

V3. The scientific literature considered for elaborating this conceptual product (ISO/IEC 29110 -Basic Profile- for BDAS +) does not present important omissions for the topic.

V4. The conceptual product (ISO/IEC 29110 -Basic Profile- for BDAS +) is logically coherent.

V5. The conceptual product (ISO/IEC 29110 -Basic Profile- for BDAS +) is adequate for achieving the purpose of its utilization.

V6. The conceptual product (ISO/IEC 29110 -Basic Profile- for BDAS +) provides new scientific-based knowledge (e.g. it is not a just a duplication of an existent conceptual product).

V7. The presentation style of the conceptual product (ISO/IEC 29110 -Basic Profile- for BDAS +) is adequate for a scientific

**Figure 7.** Median and Mean of the Seniors are represented in the 7 items of Conceptual Validity.

In addition to the quantitative results, the evaluators also provided written observations and qualitative remarks as part of the Conceptual Evaluation questionnaire. These comments offer further insights into the perceived strengths and areas for improvement of the proposed artifact, and can be reviewed in Appendix E.

### 3.8.2 Usability Evaluation of LDSAM

The LDSAM SDLC was shared with DSA academics and practitioners through the web EPG, and they were asked to evaluate usability metrics through a questionnaire based on highly cited studies (Moore & Benbasat, 1991; Karahanna et al., 1999; Lee et al., 2001). The constructs of interest used to evaluate usability of LDSAM SDLC by the panel of BDAS academics and practitioners are presented in Table 12.

A total of 27 academics and professionals from Latin America, the United States, Canada, and Europe participated in the study to obtain the usability data, which is analyzed in its entirety. The demographic data is already reported in Table 10. We provided candidates with time to read the LDSAM Methodology documentation, the EPG, and review the templates. Finally, we administered the Usability questionnaires – reported in the Appendix C –, alternating the questions: the first related to the LDSAM Methodology, and the second related to another BDAS Methodology with which the participant had experience.

**Table 12.** Constructs to be Evaluated for the Sample of International Academics and Practitioners on the LDSAM Methodology.

CONSTRUCT	ITEMS	SCALE
<b>USEFULNESS</b> – is the degree to which using the new TOOL is perceived as being better than using the current used TOOL.	4	5-points Likert (1: strongly disagree to 5: strongly agree)
<b>EASE OF USE</b> – is the degree to which using the new TOOL is perceived as being free of effort.	3	5-points Likert (1: strongly disagree to 5: strongly agree)
<b>VALUE</b> – the degree to which using the new TOOL is perceived as a value delivery entity for users by savings on money, time, and the provision of a variety of valuable resources, and by an overall value.	4	5-points Likert (1: very low to 5: very high)
<b>ATTITUDE</b> – it reflects the individual's positive and negative evaluations of performing the behavior (of adopting the evaluated artifact).	3	7-point Semantic differential scale (-3 to +3)

Statistical analyses regarding the statistical reliability and validity convergent and discriminant of the Usability instruments were performed using the Partial Least Squares (PLS) method (Gefen et al. 2018; Hair et al. 2019; Russo and Stol, 2021). Table 13 summarizes the validation criteria usually considered in the core PLS literature and implemented through the academic free version of SmartPLS v4.

PLS (Gefen et al. 2018; Hair et al. 2019; Russo and Stol, 2021) is a second-generation multivariate analysis technique that is particularly suitable for: (1) simultaneously estimating construct reliability, convergent validity, discriminant validity, and regression coefficients among the hypothesized relationships between constructs (i.e., path analysis), as well as the explained variance ( $R^2$ ) of the dependent constructs; (2) analytical scenarios involving relatively small sample sizes; and (3) situations in which the dataset does not meet the assumptions of normal distribution across constructs. Tables 14 and 15 presents, respectively, the reliability, the convergent validity, and the discriminant validity metrics for the LDSAM SDLC and the evaluated alternative BDAS SDLC, based on the collected assessment data. Reliability is reported using the Composite Reliability Index. Convergent validity is expressed through the Factor Loadings of each construct. Discriminant validity is assessed using the Average Variance Extracted (AVE).

The results in Tables 14 and 15 provide supporting evidence that the constructs USEFULNESS, EASE OF USE, VALUE, and ATTITUDE POTENTIAL USAGE were measured with satisfactory levels of reliability, convergent validity, and discriminant validity based in the usual validation criteria (Table 13). Figures 8 and 9 illustrate the SmartPLS construct diagrams used for the LDSAM SDLC and the alternative BDAS SDLC.

**Table 13.** Summary of PLS Validation Criteria.

Validation Dimension	Metric	Gefen et al. (2000) criterion	Hair et al. (2019) criterion	Russo and Stol (2021) criterion
Reliability	Composite Reliability Index (a sub-type of Internal Consistency Index)	The internal consistency index $\geq 0.700$	The internal consistency reliability index $\geq 0.700$ and $\leq 0.950$	The internal consistency reliability index $\geq 0.700$ and $\leq 0.900$
Convergent Validity	Factor Loadings of each Construct	The item loadings $\geq 0.707$	Reflective indicator loadings $\geq 0.708$  AVE $\geq 0.500$	The loading of an item $\geq 0.708$  AVE $\geq 0.500$
Discriminant Validity	Average Variance Extracted (AVE)	Each construct AVE should be larger than its correlation with other constructs.  Tables 14 and 15 (bottom part of the Table).	The shared variance for all model constructs should not be larger than their AVEs.  Tables 14 and 15 (bottom part of the Table).	All correlations between constructs be less than the lowest of the square root values of the AVEs.  Tables 14 and 15 (bottom part of the Table).

We report that the four Composite Reliability indices for the constructs USEFULNESS, EASE OF USE, VALUE, and ATTITUDE / POTENTIAL USAGE in both PLS models (LDSAM SDLC and the alternative BDAS SDLC) fell within the satisfactory range of 0.700 to 0.950 (Table 13). Regarding convergent validity, factor loadings were examined and were expected to exceed the threshold of 0.708 (Table 13). All loading values met this expectation in both PLS models, except for the items E3X (0.443) and E3Y (-0.057) in the PLS models of the LDSAM SDLC (see Figure 8) and alternative BDAS SDLC (see Figure 9). Nevertheless, the item loadings across both models were the highest within their respective rows, satisfying the expected condition for convergent validity. Discriminant validity was evaluated using the classic Fornell–Larcker criterion, which requires that the square root of the AVE (Average Variance Extracted) for each construct be greater than 0.708 (i.e.,  $AVE > 0.500$ ) and higher than its correlations with the other constructs (Table 13). The AVE values for the constructs USEFULNESS, EASE OF USE, VALUE, and ATTITUDE / POTENTIAL USAGE satisfied this criterion across both PLS models. Accordingly, discriminant validity is also reported as satisfactory for these constructs in both models. In addition to the quantitative usability metrics, evaluators were invited to provide qualitative feedback regarding their experience with the proposed methodology.

Tables 16 and 17 report that the results were favorable for the LDSAM Methodology in four Usability constructs: USEFULNESS, EASE OF USE, VALUE, and ATTITUDE. Comparing the LDSAM Methodology to another BDAS Methodology yielded the following factor results: Usefulness, the mean score was 4.213 compared to 3.519; Ease of Use, the mean score was 4.222 versus 3.519; Value, the mean score was 4.222 versus 3.111. Finally, the overall score was 1.556 compared to 1.050.

**Table 14.** Reliability, Convergent validity and discriminant validity metrics of usability constructs for the LDSAM SDLC from the 27 evaluators.

		USEFULNESS	EASE OF USE	VALUE	ATTITUDE POTENTIAL USAGE
COMPOSITE RELIABILITY INDEX >= 0.700 and <= 0.950		0.797	0.934	0.935	0.938
		FACTOR LOADINGS			
CONVERGENT VALIDITY OF CONSTRUCT (FACTOR LOADING FOR EACH ITEM >= 0.708)	EOU1	0.363	0.872	0.384	0.269
	EOU2	0.47	0.895	0.327	0.285
	EOU3	0.213	0.443	0.221	0.212
	APU1	0.761	0.424	0.793	0.948
	APU2	0.816	0.36	0.846	0.954
	APU3	0.602	0.068	0.586	0.833
	USF1	0.812	0.246	0.635	0.69
	USF2	0.884	0.416	0.617	0.715
	USF3	0.926	0.456	0.714	0.795
	USF4	0.907	0.542	0.661	0.649
	VAL1	0.602	0.122	0.799	0.713
	VAL2	0.586	0.45	0.867	0.626
	VAL3	0.748	0.446	0.901	0.776
	VAL4	0.678	0.421	0.964	0.786
		CORRELATIONS			
DISCRIMINANT VALIDITY OF CONSTRUCT (SQUARE ROOT OF AVE >= 0.708)	USEFULNESS	0.883	0.479	0.744	0.805
	EASE OF USE	0.479	0.765	0.41	0.333
	VALUE	0.744	0.41	0.885	0.825
	ATTITUDE POTENTIAL USAGE	0.805	0.333	0.825	0.913

**Table 15.** Reliability, Convergent validity and discriminant validity metrics of usability constructs for the Alternative BDAS SDLC from the 27 evaluators.

		USEFULNESS	EASE OF USE	VALUE	ATTITUDE POTENTIAL USAGE
<b>COMPOSITE RELIABILITY INDEX</b> ≥ 0.700 and ≤ 0.950		0.950	0.663	0.944	0.918
		FACTOR LOADINGS			
<b>CONVERGENT VALIDITY OF CONSTRUCT (FACTOR LOADING FOR EACH ITEM ≥ 0.708)</b>	EOU1	0.472	0.933	0.547	0.387
	EOU2	0.319	0.818	0.482	0.288
	EOU3	-0.103	-0.057	-0.379	-0.177
	APU1	0.139	-0.018	0.221	0.718
	APU2	0.585	0.389	0.542	0.977
	APU3	0.745	0.527	0.586	0.952
	USF1	0.856	0.288	0.585	0.484
	USF2	0.927	0.36	0.55	0.591
	USF3	0.943	0.419	0.651	0.587
	USF4	0.904	0.593	0.703	0.609
	VAL1	0.715	0.619	0.917	0.548
	VAL2	0.698	0.618	0.939	0.486
	VAL3	0.474	0.264	0.797	0.324
	VAL4	0.57	0.714	0.939	0.583
		CORRELATIONS			
<b>DISCRIMINANT VALIDITY OF CONSTRUCT (SQUARE ROOT OF AVE ≥ 0.708)</b>	USEFULNESS	0.908	0.474	0.693	0.63
	EASE OF USE	0.474	0.717	0.638	0.414
	VALUE	0.693	0.638	0.9	0.55
	ATTITUDE POTENTIAL USAGE	0.63	0.414	0.55	0.89

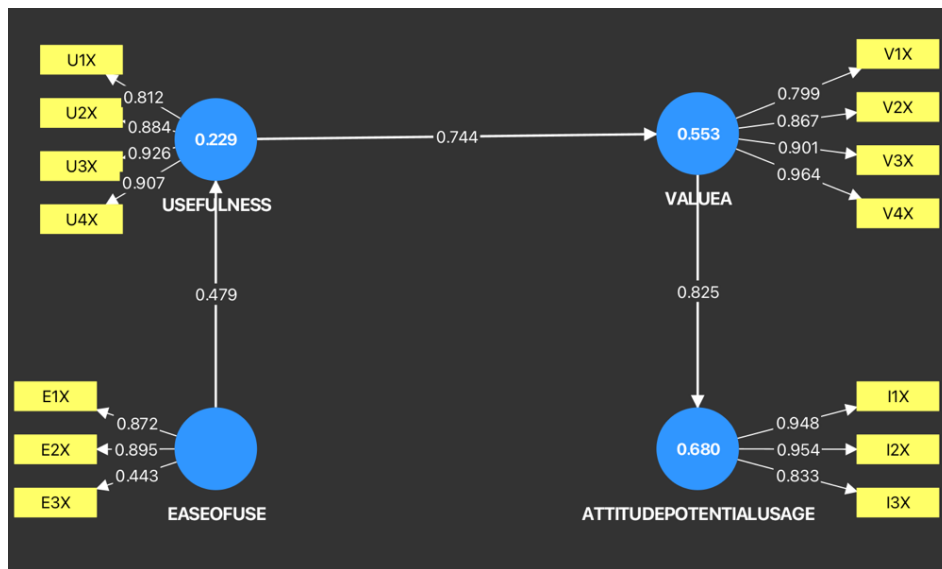


Figure 8. SDLC LDSAM Diagram in SmartPLS.

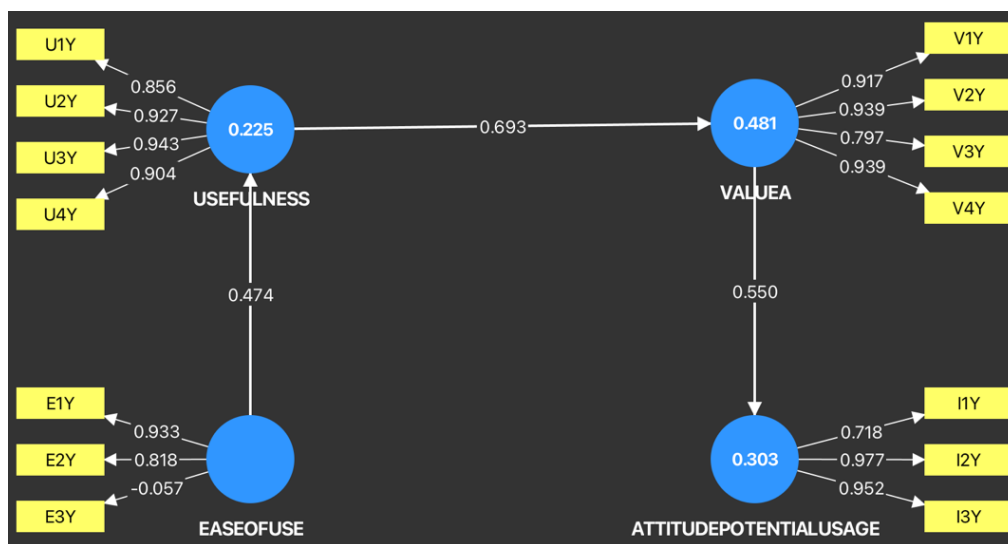


Figure 9. Alternative BDAS SDLC Diagram in SmartPLS.

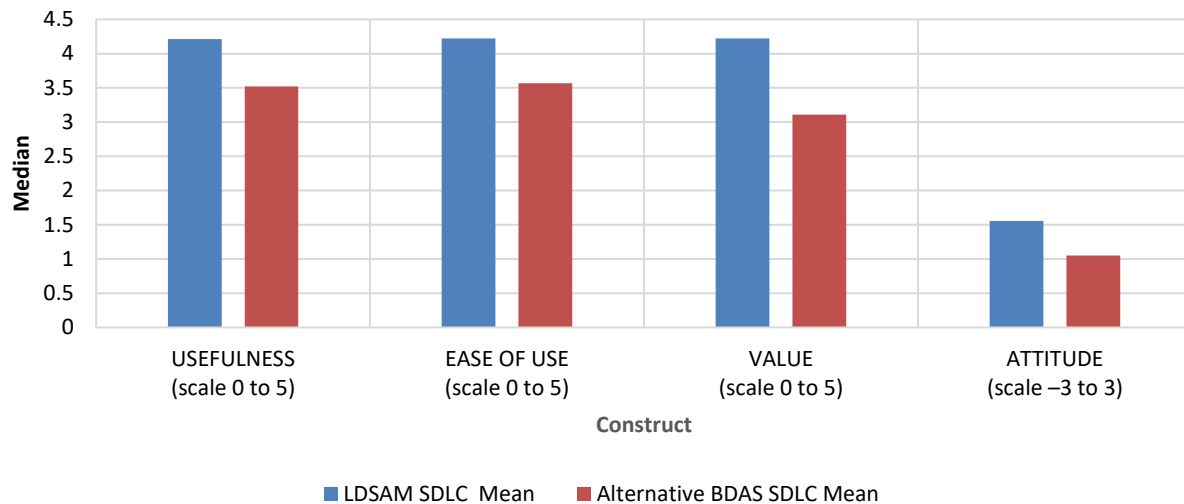
To support the statistical validity of these results, we conducted four Statistical Hypotheses Tests (SH0.1–SH0.4) reported in Table 18 by using a Paired Wilcoxon Signed-Rank Tests of medians (Sheskin, 2003). This statistical test is a non-parametric that is used when the requirement on adjustment to Normal distribution of the comparison datasets is not supported using a Shapiro-Wilk test. Table 18 and Figure 10 shows satisfactory results for the five cases, and thus we can reject all statistical null hypotheses (all p-values were less than 0.001), and support that the medians for the constructs USEFULNESS, EASE OF USE, VALUE and ATTITUDE for LDSAM Methodology – labeled as X SDLC – are greater than the ones for the alternative a BDAS Methodology considered by the 27 Evaluators – labeled as Y SDLC –.

Table 16. Usability results for LDSAM Methodology.

Construct	Median	Mean	Standard Dev
USEFULNESS	4.000	4.213	0.789
EASE OF USE	4.000	4.222	0.880
VALUE	4.000	4.222	0.765
ATTITUDE	2.000	1.556	1.095

**Table 17.** Usability results for other BDAS methodology that applicants know.

Construct	Median	Mean	Standard Dev
USEFULNESS	3.500	3.519	0.791
EASE OF USE	4.000	3.568	0.894
VALUE	3.000	3.111	0.777
ATTITUDE	1.000	1.050	1.066

**Figure 10.** Comparison of LDSAM SDLC Median and Alternative BDAS SDLC Median.**Table 18.** Wilcoxon Signed-Rank tests for the usability constructs in LDSAM vs alternative BDAS methodology.

Null Hypothesis	LDSAM SDLC Median (med.1)	Alternative BDAS SDLC Median (med.2)	P-value	Implication
SH0.1 For USEFULNESS construct (med.1 ≤ med.2)	4.000	3.500	< 0.001	SH0.1 is rejected, and thus the USEFULNESS of LDSAM SDLC is better.
SH0.2 For EASE OF USE construct (med.1 ≤ med.2)	4.000	4.000	< 0.001	SH0.2 is rejected, and thus the EASE OF USE of LDSAM SDLC is better.
SH0.3 For VALUE construct (med.1 ≤ med.2)	4.000	3.000	< 0.001	SH0.3 is rejected, and thus the VALUE of LDSAM SDLC is better.
SH0.4 For ATTITUDE construct (med.1 ≤ med.2)	2.000	1.000	0.002	SH0.4 is rejected, and thus the ATTITUDE OF POTENTIAL USAGE of LDSAM SDLC is better.

These written comments address perceived benefits of using ISO/IEC 29110 – Basic Profile – for BDAS+, as well as benefits derived from their current tool (methodology). Likewise, evaluators reported limitations associated with ISO/IEC 29110 – Basic Profile – for BDAS+, together with limitations observed in their current tool (methodology). All qualitative remarks and narrative responses have been compiled for examination in Appendix F.

## 4 DISCUSSION OF CONTRIBUTIONS AND LIMITATIONS

Regarding the research null hypotheses proposed in the Introduction: H0.1 was rejected, as the expert panel validation confirmed the existence of valid methodological elements for elaborating LDSAM (see Section 3). H0.2.1 was rejected, as LDSAM was successfully documented in an EPG. H0.2.2 was rejected, as the pilot usability evaluation yielded satisfactory results across all constructs (see Tables 16–18). Accordingly, RQ.1 is answered

affirmatively (valid elements exist and were theoretically validated) and RQ.2 is answered affirmatively, LDSAM can be documented in an EPG and was rated as useful, easy to use, and valuable by BDAS academics and practitioners.

#### 4.1 Discussion of Contributions to the Theoretical Knowledge on Development Methodologies for BDAS

This research makes several theoretical contributions to the field of BDAS development methodologies. Rather than proposing an isolated model, it systematically integrates and evaluates elements from existing Lightweight Development and BDAS frameworks to construct a novel methodological perspective. In addressing RQ.1, the study identifies and validates the theoretical components required to design a Lightweight Development Methodology for BDAS Software Systems, as confirmed by an Expert Panel.

##### Key theoretical contributions include:

- **Synthesis of Development Paradigms:** The study demonstrates that combining principles from Lightweight Development and BDAS methodologies enables the design of scalable, iterative, and adaptive frameworks. This synthesis advances current theory by revealing which components are most relevant for BDAS contexts.
- **Conceptual Model for BDAS Development:** The research proposes a structured conceptual framework that articulates key phases, activities, and integration points. This contribution does not merely restate the literature but organizes it into a coherent model that can be used as a foundation for further theory-building.
- **Empirical Support via Expert Validation:** Expert evaluation confirms the relevance and internal consistency of the proposed theoretical components, which allowed the study to reject H0.1. This validation strengthens the theoretical legitimacy of the model and increases confidence in its future application.

##### Implications for theory:

This research extends the theoretical foundation for BDAS development by identifying and organizing essential components into a cohesive model. It provides a foundation for future refinements, comparisons against emerging methodologies, and empirical testing in diverse domains.

#### 4.2 Contributions to Practice and Methodological Implementation

In addition to theoretical contributions, this research delivers a series of pragmatic outcomes that address RQ.2 by operationalizing the methodology through an EPG. These practical contributions demonstrate that the proposed model can be implemented and used by practitioners in real environments.

##### Key practical contributions include:

- **Creation of an EPG:** The methodology was successfully formalized within an EPG, demonstrating that a BDAS methodology can be represented as a structured digital resource. This finding contradicts H0.2.1.
- **Evaluation of Practical Utility:** The pilot evaluation with BDAS academics and practitioners revealed satisfactory results on usability metrics, challenging H0.2.2 and suggesting that the methodology is intuitive and usable in practice.
- **Templates to Support Adoption:** The inclusion of templates for products and deliverables supports consistent adoption and minimizes ambiguity during implementation. These resources are accessible from the corresponding author on request.
- **Bridging Theory and Practice:** The transformation of the theoretical model into an operational guide addresses a long-standing gap in BDAS methodologies: the ability to translate conceptual models into actionable processes.

##### Implications for practice:

The LDSAM framework provides organizations and research teams with a structured, lightweight approach that supports analytical rigor while aligning with the dynamics of software development. Although the templates are not prescriptive, they offer a starting point that can be adapted based on the specific needs of each project.

### 4.3 Limitations and Critical Reflections

Despite the valuable contributions of this research, several limitations require deeper consideration. These limitations contextualize the findings and delineate areas for further investigation.

#### Sample Size and Panel Composition

A key limitation relates to the size of the expert panel and participant group used during the evaluation phase. While the study gathered feedback from 27 experts and practitioners, this number is modest when compared with broader empirical standards. However, recruiting individuals with direct experience in BDAS frameworks, methodological development, and software engineering proved notably challenging. Experts in this domain are scarce and often engaged in industry roles that limit availability for research participation.

Even though the sample size is relatively small, it aligns with acceptable thresholds for exploratory methodological evaluation, especially when supported by established analytical tools such as Smart PLS. Prior studies indicate that Smart PLS is suitable for small to medium sample sizes, allowing the findings to retain statistical validity. Nevertheless, future studies would benefit from increased participant diversity to expand the representativeness of perspectives and strengthen external validity.

#### Lack of Real-World Implementation

Another limitation stems from the absence of full-scale implementation in real-world industrial environments. The current evaluation focuses on expert validation and pilot testing, which provides initial evidence of usability, structure, and conceptual soundness. However, practical deployment within ongoing BDAS projects has not yet been conducted.

Real-world implementation is expected as future work and is essential to assess:

- long-term sustainability,
- impact on project outcomes,
- adaptation to organizational constraints,
- behavior under iterative development cycles.

Such applications will allow measurement of key performance indicators (development time, error rates, model stability, and stakeholder satisfaction) providing stronger evidence of the methodology's effectiveness.

#### Contextual Constraints in Very Small Entities

The applicability of the LDSAM to VSEs is particularly relevant, given the origin of ISO/IEC 29110. However, VSE contexts introduce challenges that were not fully investigated in this study. VSEs typically operate with limited staffing, computational capacity, and process formalization. These factors may influence:

- the ability to adopt iterative analytical workflows,
- access to data engineering resources,
- feasibility of producing structured deliverables.

Initial results suggest that LDSAM may offer a promising lightweight alternative suitable for VSEs, especially those initiating BDAS development and seeking standards-based validation. Nonetheless, empirical validation within real VSE settings is necessary to determine whether:

- templates are practical without burdening teams,
- roles and activities can be effectively distributed,
- scalability can be maintained as project complexity increases.

#### Summary of Impact of Limitations

Taken together, these limitations do not invalidate the contributions of the research. Instead, they outline the boundaries within which the results should be interpreted:

- The insights gained are adequate for exploratory and methodological development stages.
- Further work is required to strengthen generalizability through larger expert groups and real-world deployments.

- VSE-specific evaluation will help determine the scope, efficiency, and scalability of LDSAM in constrained environments.
- The EPG was tested within a defined pilot group, and its broader applicability to different BDAS environments remains an area for further study.

Future studies addressing these areas will provide additional evidence to reinforce the methodology's theoretical robustness and practical usability. Presented constraints suggest that while LDSAM is a promising contribution, additional empirical studies are required, particularly longitudinal and multi-context evaluations, to determine the extent of its applicability and durability.

## 5 CONCLUSIONS

This research presents LDSAM as a distinctive contribution to the field of Big Data Science – Analytics Software development. Rather than adapting traditional SDLC models to analytical environments, LDSAM establishes a dedicated development methodology specifically designed for BDAS projects. This represents a departure from existing frameworks, which either lack analytical capabilities or introduce unnecessary complexity when applied directly to data-driven systems.

The unique value of LDSAM lies in its integration of analytical lifecycle principles within a lightweight, standards-based structure. The methodology formalizes six analytical phases, defines corresponding activities and roles, and introduces mandatory internal products, such as data cleaning reports and model evaluation artifacts, that ensure empirical evidence is generated and preserved throughout development. These features provide a governance layer that is absent in existing SDLC approaches when applied to big data and analytics environments.

In addition, LDSAM advances methodological usability by being operationalized through an EPG, enabling practitioners to adopt the methodology without requiring extensive process engineering expertise. The availability of product templates and structured guidance facilitates adoption across different organizational contexts, positioning LDSAM as a pragmatic bridge between theoretical methodological design and real-world implementation.

The study also demonstrates that LDSAM is not merely an adaptation of ISO/IEC 29110 but an extension that expands the standard to accommodate the realities of analytical systems. This extension preserves the clarity and discipline of ISO/IEC 29110 while enabling iterative analysis, model management, and evidence-based decision-making, capabilities that existing SDLC frameworks do not provide.

Future research should pursue empirical implementation in real BDAS projects, evaluate scalability across different organizational sizes including Very Small Entities, and perform longitudinal assessments to determine sustainability and impact over multiple project cycles. These efforts will provide deeper insights into the effectiveness of LDSAM and guide further refinement of both its theoretical foundations and practical tools.

## ABBREVIATIONS

Abbreviation	Full Name
AVE	Average Variance Extracted
BDAS	Big Data Analytics Systems
BDPL	Big Data Project Lifecycle
CRISP-DM	Cross-Industry Standard Process for Data Mining
DDSL	Domino Data Science Lifecycle
DSRM	Design Science Research Methodology
EPG	Electronic Process Guide
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization

Abbreviation	Full Name
LDSAM	Lightweight Data Science Analytics Methodology
PLS	Partial Least Squares
PM	Project Management
RUP	Rational Unified Process
SDLC	System Development Life Cycle
SI	Software Implementation
SPI	Software Process Improvement
TDSP	Team Data Science Process
VSE	Very Small Entity

## ADDITIONAL INFORMATION AND DECLARATIONS

**Acknowledgments:** We gratefully acknowledge the support provided by our universities in the development of this research. We also extend our sincere thanks to the professors, scholars, and domain experts whose valuable insights and expertise contributed to this work. D.A.M.M.: I would like to express my deepest gratitude to my daughter Emma, whose inspiration and energy have enabled me to reach my full potential.

**Conflict of Interests:** The authors declare no conflict of interest.

**Author Contributions:** D.A.M.M.: Conceptualization, Methodology, Software, Data curation, Formal analysis, Writing – Original draft preparation, Writing – Reviewing and Editing. M.M.: Methodology, Validation, Formal analysis, Writing – Reviewing and Editing. G.P.W.: Conceptualization, Writing – Reviewing and Editing, Supervision. S.G.C.: Investigation, Validation, Writing – Reviewing and Editing. E.L.M.A.: Data curation, Visualization, Writing – Reviewing and Editing.

**Statement on the Use of Artificial Intelligence Tools:** The authors declare that they didn't use artificial intelligence tools for text or other media generation in this article.

**Data Availability:** The data that supports the findings of this study are available from the corresponding author. Appendices A–F are available online on the article webpage: <https://doi.org/10.18267/j.aip.317>.

## REFERENCES

- Ajah, I. A., & Nweke, H. F. (2019). Big data and business analytics: trends, platforms, success factors and applications. *Big Data and Cognitive Computing*, 3(2), 32. <https://doi.org/10.3390/bdcc3020032>
- Beecham, S., Hall, T., Britton, C., Cottee, M., & Rainer, A. (2005). Using an expert panel to validate a requirements process improvement model. *Journal of Systems and Software*, 76(3), 251–275. <https://doi.org/10.1016/j.jss.2004.06.004>
- Beck, K. (1999). Embracing change with extreme programming. *Computer*, 32(10), 70–77. <https://doi.org/10.1109/2.796139>
- Blalock, H. M. (1969). *Theory construction: From verbal to mathematical formulations*. Prentice-Hall.
- Boehm, B., & Turner, R. (2003). Using risk to balance agile and plan-driven methods. *Computer*, 36(6), 57–66. <https://doi.org/10.1109/MC.2003.1204376>
- Chapman, Clinton J., Kerber R., Khabaza T., Reinartz T., Shearer C., & Wirth R. (2000). *CRISP-DM 1.0 step-by-step data mining guide*. SPSS. <https://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/es/CRISP-DM.pdf>
- Clarke, P., & O'Connor, R. V. (2013). An empirical examination of the extent of software process improvement in software SMEs. *Journal of Software Evolution and Process*, 25(9), 981–998. <https://doi.org/10.1002/smr.1580>
- Conboy, K. (2009). Agility from First Principles: Reconstructing the Concept of Agility in Information Systems Development. *Information Systems Research*, 20(3), 329–354. <https://doi.org/10.1287/isre.1090.0236>
- Davoudian, A., & Liu, M. (2020). Big data systems: A software engineering perspective. *ACM Computing Surveys*, 53(5), 1–39. <https://doi.org/10.1145/3408314>
- Domino Data Lab. (2019). The Practical Guide to Managing Data Science at Scale. Domino Data Lab. <https://domino.ai/resources/managing-data-science>
- Karahanna, E., Straub, D. W., & Chervany, N. L. (1999). Information Technology Adoption Across Time: A Cross-Sectional Comparison of Pre-Adoption and Post-Adoption Beliefs1. *MIS Quarterly*, 23(2), 183–213. <https://doi.org/10.2307/249751>

- Ebert, C. (2007). The impacts of software product management. *Journal of Systems and Software*, 80(6), 850–861. <https://doi.org/10.1016/j.jss.2006.09.017>
- Fayyad, U. M., Haussler, D., & Stolorz, P. E. (1996). KDD for Science Data Analysis: Issues and Examples. In *KDD-96 Proceedings*, (pp. 50–56). AAAI. <https://cdn.aaai.org/KDD/1996/KDD96-009.pdf>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11), 27–34. <https://doi.org/10.1145/240455.240464>
- Galvan-Cruz, S., Mora, M., Laporte, C. Y., & Duran-Limon, H. (2021). Reconciliation of scrum and the project management process of the ISO/IEC 29110 standard-Entry profile—an experimental evaluation through usability measures. *Software Quality Journal*, 29(2), 239–273. <https://doi.org/10.1007/s11219-021-09552-3>
- Gefen, D., Straub, D., & Boudreau, M. (2000). Structural Equation Modeling and Regression: Guidelines for Research practice. *Communications of the Association for Information Systems*, 4, Article 7. <https://doi.org/10.17705/1cais.00407>
- Gornik, D. (2003). Rational Unified Process: Best Practices for Software Development Teams. IBM. [https://public.dhe.ibm.com/software/rational/web/whitepapers/2003/rup\\_bestpractices.pdf](https://public.dhe.ibm.com/software/rational/web/whitepapers/2003/rup_bestpractices.pdf)
- Gorton, I., Bener, A. B., & Mockus, A. (2016). Software engineering for big data systems. *IEEE Software*, 33(2), 32–35. <https://doi.org/10.1109/ms.2016.47>
- Greeno, G. J., Simon, & A., H. (1988). Problem Solving and Reasoning. In *Handbook of Experimental Psychology*, (pp. 589–671). John Wiley & Sons.
- Grunstra, B. R., Ackoff, R. L., Gupta, S. K., & Minas, J. S. (1965). Scientific method: Optimizing applied research decisions. *Philosophy and Phenomenological Research*, 25(4), 594. <https://doi.org/10.2307/2105451>
- Hakim, A., Indonesia, U. B., Suryadi, N., Huda, C., Indonesia, U. B., & Indonesia, U. B. (2024). Sharia Retail store service standards based on customer preferences in the cooperative ecosystem. *Jurnal Aplikasi Manajemen*, 22(2), 335–363. <https://doi.org/10.21776/ub.jam.2024.022.02.05>
- Haakman, M., Cruz, L., Huijgens, H., & van Deursen, A. (2021). AI lifecycle models need to be revised: An exploratory study in Fintech. *Empirical Software Engineering*, 26, 1–29. <https://doi.org/10.1007/s10664-021-09993-1>
- Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2–24. <https://doi.org/10.1108/EBR-11-2018-0203>
- Hevner, A. R. (2007). A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, 19(2), Article 4.
- Hoda, R., Salleh, N., & Grundy, J. (2018). The rise and evolution of agile software development. *IEEE Software*, 35(5), 58–63. <https://doi.org/10.1109/ms.2018.29011318>
- Hummel, O., Eichelberger, H., Giloj, A., Werle, D., & Schmid, K. (2018). A collection of software engineering challenges for big data system development. In *2018 44th Euromicro Conference on Software Engineering and Advanced Applications*, (pp. 362–369). IEEE. <https://doi.org/10.1109/SEAA.2018.00066>
- Iranmanesh, M., Lim, K. H., Foroughi, B., Hong, M. C., & Ghobakhloo, M. (2023). Determinants of intention to adopt big data and outsourcing among SMEs: organisational and technological factors as moderators. *Management Decision*, 61(1), 201–222. <https://doi.org/10.1108/md-08-2021-1059>
- ISO/IEC. (2011). ISO/IEC TR 29110-5-1-2:2011 – Software Engineering – Lifecycle Profiles for Very Small Entities (VSEs) – Part 5-1-2: Management and engineering guide – Generic profile group: Basic profile. International Organization for Standardization/International Electrotechnical Commission. <https://www.iso.org/es/contents/data/standard/05/11/51153.html>
- ISO/IEC. (2023). Systems and Software Engineering – System Life Cycle Processes ISO/IEC15288:2023. International Organization for Standardization/International Electrotechnical Commission. <https://www.iso.org/standard/81702.html>
- ISO/IEC. (2017). Systems and Software Engineering – Software Life Cycle Processes, ISO/IEC/IEEE 12207:2026. International Organization for Standardization/International Electrotechnical Commission. <https://www.iso.org/standard/63712.html>
- Kitchin, R., & Lauriault, T. P. (2015). Small data in the era of big data. *GeoJournal*, 80(4), 463–475. <https://doi.org/10.1007/s10708-014-9601-7>
- Kose, B.O. (2021). Agile business analysis for digital transformation. In *Handbook of Research on Multidisciplinary Approaches to Entrepreneurship, Innovation, and ICTs*, (pp. 98–123). IGI Global. <https://doi.org/10.4018/978-1-7998-4099-2.ch006>
- Laporte, C., O'Connor, R., & Fanmuy, G. (2013) International systems and software engineering standards for very small entities. *CrossTalk*, 2013(May/June), 28–33.
- Lee, D., Park, J., & Ahn, J.-H. (2001). On the Explanation of Factors Affecting E-Commerce Adoption. In *ICIS 2001 Proceedings International Conference on Information Systems*, (Article 14). AIS.
- Lin, Y., & Huang, S. (2018). The design of a software engineering lifecycle process for big data projects. *IT Professional*, 20(1), 45–52. <https://doi.org/10.1109/mitp.2018.011291352>
- Mariscal, G., Marbán, Ó., & Fernández, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review*, 25(2), 137–166. <https://doi.org/10.1017/s0269888910000032>
- Maroufkhani, P., Tseng, M., Iranmanesh, M., Ismail, W. K. W., & Khalid, H. (2020). Big data analytics adoption: Determinants and performances among small to medium-sized enterprises. *International Journal of Information Management*, 54, 102190. <https://doi.org/10.1016/j.ijinfomgt.2020.102190>
- Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2019). CRISP-DM Twenty years Later: From data mining processes to data science trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), 3048–3061. <https://doi.org/10.1109/tkde.2019.2962680>

- Martinez, I., Viles, E., & Olaiola, I. G. (2021). Data Science Methodologies: Current challenges and Future Approaches. *Big Data Research*, 24, 100183. <https://doi.org/10.1016/j.bdr.2020.100183>
- Microsoft. (2017) Team Data Science Process Lifecycle. <https://github.com/Azure/Microsoft-TDSP/blob/master/Docs/lifecycle-detail.md>
- Mohd-Selamat, S. A., Prakoonwit, S., Sahandi, R., Khan, W., & Ramachandran, M. (2018). Big data analytics—A review of data-mining models for small and medium enterprises in the transportation sector. *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery*, 8(3), e1238. <https://doi.org/10.1002/widm.1238>
- Montoya-Murillo, D., Mora, M., Galvan-Cruz, S., Munoz-Zavala, A., & Alvarez-Rodriguez, F. (2025). A Review of SDLCs for Big Data Analytics Systems in the Context of Very Small Entities Using the ISO/IEC 29110 Standard-Basic Profile. *International Arab Journal of Information Technology*, 22(1), 194–215. <https://doi.org/10.34028/iajit/22/1/15>
- Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, 2(3), 192–222. <https://doi.org/10.1287/isre.2.3.192>
- Mora, M., Gelman, O., Paradice, D., & Cervantes, F. (2008). The case for conceptual research in information systems. In *CONF-IRM 2008 Proceedings International Conference on Information Resources Management*, (Article 52). AIS.
- Mora, M., Raisinghani, M. S., O'Connor, R., & Gelman, O. (2009). Toward an integrated conceptualization of the service and service system concepts. *International Journal of Information Systems in the Service Sector*, 1(2), 36–57. <https://doi.org/10.4018/jisss.2009040103>
- Newell, A., & Simon, H. A. (2019). *Human problem solving*. Echo Point Books & Media.
- Niazi, M. (2015). A comparative study of software process improvement implementation success factors. *Journal of Software Evolution and Process*, 27(9), 700–722. <https://doi.org/10.1002/smr.1704>
- O'Connor, R. V., & Laporte, C. Y. (2017). The evolution of the ISO/IEC 29110 set of standards and guides. *International Journal of Information Technologies and Systems Approach*, 10(1), 1–21. <https://doi.org/10.4018/ijitsa.2017010101>
- O'Connor, R. V., & Coleman, G. (2009). Ignoring “Best Practice”: Why Irish Software SMEs are Rejecting CMMI and ISO 9000. *Australian Journal of Information Systems*, 16(1), 7–30. <https://doi.org/10.3127/ajis.v16i1.557>
- Oktaba, H., & Ibarguengoitia González, G. (1998). Software process modeled with objects: Static view. *Computación y Sistemas*, 1(4), 228–238.
- Olorunshola, O. E., & Ogwueleka, F. N. (2021). Review of system development life cycle (SDLC) models for effective application delivery. In *Information and Communication Technology for Competitive Strategies (ICTCS 2020) ICT: Applications and Social Interfaces* (pp. 281–289). Springer. [https://doi.org/10.1007/978-981-16-0739-4\\_28](https://doi.org/10.1007/978-981-16-0739-4_28)
- Pääkkönen, P., & Pakkala, D. (2015). Reference architecture and classification of technologies, products and services for big data systems. *Big Data Research*, 2(4), 166–186. <https://doi.org/10.1016/j.bdr.2015.01.001>
- Pai, D. R., Subramanian, G. H., & Pendharkar, P. C. (2015). Benchmarking software development productivity of CMMI level 5 projects. *Information Technology and Management*, 16(3), 235–251. <https://doi.org/10.1007/s10799-015-0234-4>
- Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Plotnikova, V., Dumas, M., & Milani, F. (2020). Adaptations of data mining methodologies: a systematic literature review. *PeerJ Computer Science*, 6, e267. <https://doi.org/10.7717/peerj-cs.267>
- Qumer, A., & Henderson-Sellers, B. (2008). An Evaluation of the Degree of Agility in Six Agile Methods and its Applicability for Method Engineering. *Information and Software Technology*, 50, 4, 280–295. <https://doi.org/10.1016/j.infsof.2007.02.002>
- Robinson, S., & Gillis, A. S. (2025). 5V's of big data. TechTarget.com. <https://www.techtarget.com/searchdatamanagement/definition/5-Vs-of-big-data>
- Russo, D., & Stol, K. (2021). PLS-SEM for Software Engineering Research. *ACM Computing Surveys*, 54(4), 1–38. <https://doi.org/10.1145/3447580>
- Saadatmand, M. (2024). A hierarchical decision model for evaluating the Strategy Readiness of Quantitative Machine Learning/Data science-driven investment strategies. Doctoral dissertation. Portland State University. <https://doi.org/10.15760/etd.3744>
- Salazar-Salazar, G., Mora, M., Duran-Limon, H., Alvarez-Rodriguez, F., & Munoz-Zavala, A. (2024). Review of Agile SDLC for Big Data Analytics Systems in the Context of Small Organizations Using Scrum-XP. *International Arab Journal of Information Technology*, 21(6), 1089–1110. <https://doi.org/10.34028/iajit/21/6/12>
- Saltz, J. S., & Krasteva, I. (2022). Current approaches for executing big data science projects—a systematic literature review. *PeerJ Computer Science*, 8, e862. <https://doi.org/10.7717/peerj-cs.862>
- SAS. (2025). Semma data mining methodology. <http://www.sas.com/technologies/analytics/datamining/miner/semma.html>
- Shafique, U., & Qaiser, H. (2014). A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, 12(1), 217–222.
- Sheskin, D. J. (2003). *Handbook of Parametric and Nonparametric Statistical Procedures*. Chapman and Hall/CRC.
- Staples, M., Niazi, M., Jeffery, R., Abrahams, A., Byatt, P., & Murphy, R. (2007). An exploratory study of why organizations do not adopt CMMI. *Journal of Systems and Software*, 80(6), 883–895. <https://doi.org/10.1016/j.jss.2006.09.008>
- Sutherland, J. (2010). *Jeff Sutherland's Scrum Handbook*. The Scrum Training Institute Press.
- Todman, L. C., Bush, A., & Hood, A. S. (2023). 'Small Data' for big insights in ecology. *Trends in Ecology & Evolution*, 38(7), 615–622. <https://doi.org/10.1016/j.tree.2023.01.015>
- Tsai, C. W., Lai, C. F., Chao, H. C., & Vasilakos, A. V. (2015). Big data analytics: A survey. *Journal of Big Data*, 2(1), 1–32. <https://doi.org/10.1186/s40537-015-0030-3>

- 
- Vinzi, E., V., Chin, W., Henseler, J., & Wang, H. (2010). *Handbook of partial least squares: Concepts, methods and applications*. Springer. <https://doi.org/10.1007/978-3-540-32827-8>
- Wang, J., Xu, C., Zhang, J., & Zhong, R. (2022). Big data analytics for intelligent manufacturing systems: A review. *Journal of Manufacturing Systems*, 62, 738–752. <https://doi.org/10.1016/j.jmsy.2021.03.005>
- 

Acta Informatica Pragensia is published by the Prague University of Economics and Business, Czech Republic | eISSN: 1805-4951

---