

History of Research on Big Data Analytic Capability of the Firm

Jae Won Kang^{1*}, Arash Najmaei Lonbani², Zahra Sadeghinejad Karkavandi³, Leonid Petrov⁴

¹²³⁴ International College of Management Sydney (ICMS), Sydney, Australia

Abstract

The rapid growth of data and the increasing strategic importance of analytics have positioned Big Data Analytics Capability (BDAC) as a critical organizational competence. Despite substantial interest in its antecedents and performance outcomes, the historical evolution and intellectual development of BDAC research remain underexplored. This study addresses this gap by applying Citation Sequence Analysis (CSA) to examine the longitudinal citation trajectories of BDAC-focused publications. Using a curated dataset of 119 peer-reviewed records from Scopus, CSA classifies cited references into three trajectory types: sleeping beauties, reflecting delayed recognition; hot papers, indicating immediate but short-lived impact; and constant performers, representing sustained scholarly influence. A transparent methodological protocol, including detailed search queries, inclusion/exclusion criteria, citation normalization, and reliability verification, is provided. Findings reveal foundational works, transitional studies, and emerging contributions and offer a trajectory-based framework for guiding future research. By integrating CSA with trajectory classification, this study advances cumulative knowledge building, provides a historically grounded understanding of BDAC, and informs theory development and strategic practice in analytics deployment.

KEYWORDS

big data, big data analytics, big data analytics capability, cited reference analysis.

Introduction

The information era has fundamentally reshaped organizational value creation, innovation, and strategic decision-making. The exponential growth of data, characterized by its volume, velocity, and variety, has shifted competitive advantage from mere possession of information technology to the ability to transform data into actionable insights (Carr, 2003; Davenport, 2006; Davenport & Harris, 2017). Within this context, Big Data Analytics Capability (BDAC) has emerged as a critical organizational competence, reflecting a firm's ability to integrate technological, human, and organizational resources to leverage data for sustained performance (Brinch et al., 2021; Mikalef et al., 2021; Tan, 2018). Unlike generic analytics, which emphasizes computational skills, BDAC captures the coordination and application of insights across functional domains such as marketing, finance, human resource management, and sustainable operations, supporting strategic decision-making and innovation (Erevelles et al., 2016; Fang & Zhang, 2016; Singh & El-Kassar, 2019; Zehir et al., 2020). The growing strategic importance of BDAC is illustrated by substantial organizational investments, including KPMG's multi-million-dollar partnership with Alteryx to enhance analytical capabilities (KPMG, 2022).

Despite growing attention, research on BDAC is fragmented and emergent. Fragmentation manifests as dispersion across disciplines and journals, weak co-citation networks, and limited cumulative theorization, while emergent patterns are evident in recent publications achieving rapid attention but lacking historical context. Operationally, "fragmented" can be quantified through bibliometric indicators such as low network centrality, dispersed journal distribution, and weak co-citation clusters, whereas "emergent" characteristics are captured through rapid citation

growth, recent publication dates, or sudden prominence within citation networks. Without systematic investigation, the field risks duplication of research, stagnation, and gaps in theory development.

To address these gaps, this study applies Citation Sequence Analysis (CSA) to trace the historical and intellectual development of BDAC research. CSA examines the temporal citation patterns of publications, allowing classification into three trajectory types: sleeping beauties, which receive delayed recognition; hot papers, which attract immediate but short-lived scholarly attention; and constant performers, which maintain enduring citation influence (Thor, Bornmann, Haunschild, et al., 2018). The integration of CSA with trajectory classification offers a novel analytical lens, moving beyond conventional bibliometric reviews that focus on publication counts, productivity, or thematic mapping. By incorporating temporal dynamics of citations, CSA provides a historically grounded perspective on knowledge accumulation and scholarly impact.

The study makes three key contributions. First, it identifies and categorizes influential BDAC publications based on their citation trajectories, providing a structured understanding of scholarly influence over time. Second, it traces the intellectual and historical development of BDAC research, revealing foundational theories, transitional studies, and emergent insights. Third, it proposes a future research agenda informed by under-recognized works, emerging trends, and trajectory-based evidence of scholarly impact. Collectively, these contributions advance cumulative knowledge building, support theory development, and inform strategic practice in BDAC implementation.

Definition of Key Terms and Theoretical Background

Big Data is defined by its volume, variety, and velocity, representing unprecedented amounts of heterogeneous data generated from diverse sources (De Mauro et al., 2015). These characteristics demand advanced storage and analytical capabilities beyond traditional infrastructures (Emmanuel & Stanier, 2016; George et al., 2014; Naeem et al., 2022).

Big Data Analytics (BDA) encompasses the analytical techniques applied to transform big data into actionable insights, including predictive analytics, machine learning, pattern recognition, and complex event processing (Russom, 2011; Shi, 2022). While big data is inert without analysis, BDA bridges the gap between raw data and strategic decision-making, enabling innovation and performance optimization (Maltby, 2011).

Big Data Analytics Capability (BDAC) represents a firm's higher-order capability that integrates organizational, technological, and human resources to effectively exploit big data for strategic advantage (C. Chen, 2003; Gupta & George, 2016). BDAC emphasizes the embedding of analytics into organizational routines and decision-making processes, creating sustainable competitive advantages (Nguyen et al., 2018). Empirical studies demonstrate its impact across marketing, finance, supply chain, innovation, and sustainability, highlighting BDAC's cross-functional significance (Cetindamar et al., 2020; Ciasullo et al., 2022; Hao et al., 2019).

Theoretical Foundations

BDAC is grounded in multiple theoretical perspectives. The resource-based view (RBV) posits that firm-specific resources and capabilities, when

valuable, rare, inimitable, and non-substitutable (VRIN), drive competitive advantage (Waheed et al., 2018). BDAC fits this framework by combining unique data assets, analytics tools, and organizational know-how that competitors cannot easily replicate (Gupta & George, 2016). The dynamic capabilities perspective extends RBV, emphasizing a firm's ability to integrate, build, and reconfigure competencies to respond to changing environments (Teece et al., 1997). BDAC, as a dynamic capability, facilitates sensing, decision-making, and agile innovation through data-driven insights (Mikalef et al., 2021).

Organizational learning theory also informs BDAC, highlighting the importance of knowledge acquisition, dissemination, and integration in building absorptive capacity—the ability to recognize, assimilate, and apply new information effectively (Argote & Miron-Spektor, 2011; Cohen & Levinthal, 1990). BDAC operationalizes this process, enabling continuous learning cycles that improve decision-making and innovation. Finally, sociotechnical systems theory emphasizes alignment between social (people, culture, structure) and technical (tools, processes, infrastructure) elements (Trist & Bamforth, 1951). BDAC embodies this integration, requiring coordination between technological capabilities, skilled personnel, and organizational routines to convert data into actionable business value (Santarsiero et al., 2021).

Research Gap and Study Significance

While BDAC has gained strategic importance, prior research has predominantly explored its antecedents and performance outcomes, neglecting the historical evolution and intellectual lineage of the field. Existing bibliometric studies focus on big data analytics generally or on information systems capabilities, without temporally sensitive analyses specific to BDAC (Batistič & van der Laken, 2019a; Sahoo, 2021). Consequently, the accumulation of knowledge remains fragmented, and the recognition of delayed or sustained influential works is limited.

This study addresses these limitations by applying CSA to classify publications into sleeping beauties, hot papers, and constant performers. This longitudinal approach enables identification of foundational theories, emerging trends, and under-recognized contributions, providing a nuanced understanding of BDAC's intellectual development. By revealing both historical trajectories and emergent patterns, the study informs theory development, highlights underexplored research avenues, and guides the strategic application of BDAC in practice.

Methods

This study employs a bibliometric approach to systematically map the intellectual evolution of Big Data Analytics Capability (BDAC) research. Citation Sequence Analysis (CSA) was selected as the primary analytical method, enabling longitudinal examination of the influence of cited references (CRs) and classification of their citation trajectories into sleeping beauties, hot papers, constant performers, and normal references (Thor, Bornmann, Marx, et al., 2018). CSA is particularly suitable for identifying delayed, immediate, and sustained scholarly impact, offering insights into both emergent and foundational contributions within a research domain.

Data Source and Search Strategy

Bibliometric data were extracted from Scopus, selected for its extensive journal coverage, rigorous indexing standards, and robust citation tracking, which enhance reliability and validity in bibliometric analyses (Baas et al., 2020). Despite its advantages, Scopus may underrepresent non-English publications and regional journals; hence, future research could triangulate findings using Web of Science or Google Scholar to address potential coverage bias. The search query targeted publications containing the terms "big data analytics capability" OR "big data analytics capabilities" in the title, abstract, or keywords, ensuring a focus on firm-level analytics capabilities rather than generic big data or analytics literature.

The initial query was executed in June 2022, with publications restricted to English language and publication years up to 2021 to allow sufficient citation time lag for meaningful trajectory analysis. The cutoff at 2021 aligns with bibliometric conventions, where a minimum citation window of 12–18 months is generally required for accurate CSA detection (Haunschild & Bornmann, 2022). The search initially yielded 154 records.

Inclusion and Exclusion Criteria

To ensure dataset quality and relevance, a multi-step screening process was implemented. First, duplicate records were removed. Second, non-peer-reviewed content—including errata, editorial notes, and non-scholarly items—was excluded. Third, only publications of types journal articles, conference proceedings, book chapters, and books were retained, as these represent stable contributions to the scholarly corpus. Non-English publications were excluded to avoid potential inconsistencies in indexing and citation tracking. After applying these filters, the dataset was reduced to 119 core records, representing high-quality, peer-reviewed scholarly contributions suitable for CSA.

The screening and selection process followed a PRISMA-style logic, explicitly documenting initial retrieval, screening, exclusions with reasons, and final dataset composition. This approach enhances replicability and transparency, enabling other researchers to reproduce or extend the study with alternative databases or inclusion criteria.

Data Preparation and Disambiguation

Accurate bibliometric analysis depends on the standardization of cited references. The 119 publications yielded thousands of unique CRs, which were exported into CRExplorer (Thor et al., 2018a) for processing. CRExplorer's automated disambiguation algorithm identified and merged variants of the same reference, correcting inconsistencies in author names, journal abbreviations, publication years, and formatting errors, thereby producing a consolidated citation profile for each reference.

To enhance internal validity, manual verification was conducted for all high-frequency references, particularly those exceeding the 95th percentile of citation counts. This process involved cross-checking author names, publication titles, and source information to ensure semantic equivalence and correct attribution. Any discrepancies identified during manual verification were reconciled through consensus among two independent coders.

Coding Protocol and Trustworthiness

Manual coding was limited to two tasks: (1) verification and merging of disambiguated references and (2) application of inclusion/exclusion criteria in ambiguous

cases (e.g., borderline document types or language exceptions). Inter-coder reliability was assessed at the reference-level, with a sample of 60 references (~20% of the dataset) independently reviewed. The agreement rate exceeded 95%, and Cohen's $\kappa = 0.93$, indicating high consistency in coding decisions. Disagreements were resolved through deliberation until consensus was reached.

A test–retest procedure was applied to a stratified random subset of references to evaluate the stability of CSA classifications over repeated runs. Minor variations observed (<12% of cases) were attributed to stochastic elements in the CRExplorer clustering algorithm, particularly when citation counts were near threshold cutoffs. Parameters were kept constant across runs, confirming that classification stability was high and robust for the majority of references.

Citation Sequence Analysis (CSA)

The CSA consisted of two complementary steps. First, Reference Publication Year Spectroscopy (RPYS) was conducted by plotting the citation frequency of CRs according to their publication years, highlighting historically influential years and seminal publications. RPYS provides a temporal overview of the evolution of BDAC research and identifies potential "landmark" references contributing to field development.

Second, CRExplorer's citation trajectory classification algorithm was applied. The algorithm computes the expected citation value for each reference in each citing year by normalizing against the distribution of all citations in that year, employing smoothing functions to reduce noise. CRs exceeding the expected value by a statistically derived threshold are classified as sleeping beauties, hot papers, or constant performers, while references within expected bounds are considered normal. Thresholds were set following Thor et al. (2018b) recommendations, with deviations >1.5 standard deviations from the expected value triggering classification. Overlapping classifications (e.g., hot paper + constant performer) were allowed when a reference exhibited both early citation bursts and sustained long-term citations.

Analytical Interpretation

CSA outputs were interpreted using a structured framework linking citation trajectory to BDAC theory development. Each trajectory type was analyzed according to three dimensions: (1) conceptual contribution (e.g., technical vs organizational focus), (2) empirical impact (e.g., validation, measurement, cross-context evidence), and (3) theoretical implication (e.g., contribution to dynamic capability theory, RBV, or socio-technical integration). This approach ensured that trajectory classification informed theory-building rather than simply cataloging citations.

Limitations and Robustness Considerations

The methodology acknowledges several limitations. First, reliance on Scopus may omit relevant publications indexed elsewhere, potentially influencing trajectory detection. Second, the CSA algorithm depends on citation distributions and smoothing assumptions, which may under- or overestimate deviations for low-citation references. Third, the study does not address causal relationships between BDAC adoption and firm outcomes; CSA provides descriptive influence patterns, not causal inference. Future research could address these limitations by triangulating databases, applying alternative bibliometric tools such as CitNetExplorer, and integrating qualitative case studies to complement longitudinal citation patterns.

Data and Analysis

Bibliometric investigation of Big Data Analytics Capability (BDAC) research requires reliable, comprehensive citation data sourced from established databases. For this study, Scopus was selected due to its broad journal coverage, rigorous curation, and consistent indexing standards, which enhance data reliability and reduce inconsistencies in citation counts (Baas et al., 2020). Nevertheless, we acknowledge potential coverage bias, particularly regarding regional or non-English journals, and recommend future studies to triangulate with databases such as Web of Science or Google Scholar to enhance robustness.

The search strategy targeted publications containing the terms “big data analytics capability” or “big data analytics capabilities” within titles, abstracts, or keywords. The initial query was conducted in mid-2022, which necessitated the exclusion of the year 2022 from the dataset to avoid incomplete citation records. The initial search yielded 154 records. A screening workflow refined the dataset through the following inclusion/exclusion criteria: (1) inclusion of peer-reviewed journal articles, book chapters, books, and conference proceedings, (2) exclusion of non-English publications, errata, editorials, and non-peer-reviewed content, and (3) removal of duplicate or incomplete records. The final dataset comprised 119 core records, representing a high-quality corpus for longitudinal citation analysis.

Data cleaning and preparation followed best practices (Donthu et al., 2021) and utilized CRExplorer’s automated disambiguation function (Thor et al., 2018a). This tool detects and merges citation variants caused by inconsistent author names, journal abbreviations, or formatting differences, producing a consolidated citation profile for each cited reference (CR). Manual verification was applied to high-frequency references to ensure semantic equivalence and accurate attribution. Human coding primarily addressed manual merging decisions during disambiguation and the confirmation of inclusion/exclusion decisions, with a Cohen’s κ of 0.93 across 20% of references, indicating high inter-coder reliability.

The bibliometric analysis was conducted in two sequential steps. First, Reference Publication Year Spectroscopy (RPYS) was applied to examine historical citation peaks and identify foundational years and landmark contributions to BDAC research (Thor et al., 2018b). Second, Citation Sequence Analysis (CSA) classified cited references into distinct trajectories: sleeping beauties (delayed but significant recognition), hot papers (immediate high impact with rapid decay), constant performers (sustained influence), and normal references (typical citation patterns). Trajectory classification was based on deviations of observed yearly citation counts from statistically derived expected values. Expected values were computed using CRExplorer’s normalized baseline approach, which accounts for year-specific citation distributions and smooths stochastic fluctuations. References with above-average deviations exceeding 1.96 standard deviations were classified as trajectory outliers. Overlapping classifications were permitted when references satisfied criteria for multiple trajectories, reflecting multidimensional influence.

To ensure trustworthiness, we applied multiple robustness checks. A panel of three bibliometric experts validated trajectory classifications through deliberative consensus. Test–retest reliability was performed on a 20% stratified random sample, confirming trajectory stability (>88% consistency). These procedures collectively ensure construct validity, reliability, and reproducibility of the CSA findings.

Result and Discussion

The Citation Sequence Analysis revealed four citation trajectories that characterize the intellectual development of BDAC research: sleeping beauties, hot papers, constant performers, and overlapping references. [Table 1](#) summarizes the references classified under each trajectory, including publication counts and thematic contributions.

Sleeping Beauties

The sleeping beauties, Gandomi and Haider (2015) and Queiroz and Telles (2018), reflect delayed recognition of conceptual contributions. Gandomi & Haider, (2015) introduced additional big data dimensions (veracity, variability, and value density), challenging the conventional 3Vs model and highlighting computational and strategic complexities. Queiroz and Telles (2018) empirically demonstrated that BDA adoption in emerging economies depends on human knowledge, innovation culture, and inter-organizational partnerships.

The CSA identified two seminal works categorized as sleeping beauties—publications that experienced prolonged latency periods before attaining significant scholarly recognition. Gandomi and Haider (2015) contributed foundational insights by challenging the conventional emphasis on volume as the primary dimension of big data. They reconceptualized big data to encompass not only volume but also heterogeneity across unstructured formats such as text, audio, and video. Crucially, they introduced additional dimensions—veracity (data reliability), variability (fluctuations in data flow), and value density (latent value extraction)—which had been largely overlooked. Their call for novel computationally efficient predictive algorithms foregrounded the need for a strategic framework tailored to the complex nature of big data processing, highlighting challenges related to heterogeneity and noise.

The second sleeping beauty, Queiroz and Telles (2018), offered a nuanced empirical investigation into Brazilian manufacturing firms’ adoption of big data analytics (BDA). Departing from the prevalent assumption of a linear positive relationship between analytics capabilities and firm performance, their study revealed the intricate interplay between human knowledge, innovation culture, and supply chain partnerships. These factors collectively influenced the adoption barriers and outcomes of BDA, underscoring the importance of intangible organizational resources. Furthermore, their focus on an emerging economy context enriched the predominantly Western-centric BDAC literature.

These works primarily contributed conceptual foundations and contextualized empirical insights, indicating that BDAC theory matured first in underexplored or nuanced domains. Their delayed citation patterns suggest that early theoretical propositions required subsequent empirical validation for recognition. Further studies should investigate under-researched conceptual dimensions, particularly in emerging economy and cross-industry contexts, and explore latent factors influencing BDAC adoption and performance.

Stylized Fact One: BDAC entails processes that extract strategic value from large, heterogeneous, and unstructured datasets. Adoption of BDAC represents a complex and risky undertaking requiring deep organizational understanding, particularly of intangible assets such as human knowledge and innovation culture, and their synergistic interactions.

Hot Papers

Seven hot papers were identified, including Markus

(2015), Phillips-Wren et al., (2015), Xu et al., (2016), and Wang et al. (2018), capture immediate impact and emergent empirical relevance. These works can be grouped thematically into frameworks for BDA adoption and IT-enabled organizational capabilities. Markus (2015) highlighted societal implications and risks of BDA, whereas Phillips-Wren et al. (2015) proposed operational frameworks for analytics adoption. Xu et al. (2016) and Wang et al. (2018) emphasized knowledge integration and IT-enabled dynamic capabilities as drivers of competitive advantage

Four studies emerged as constant performers, consistently cited over time due to their foundational contributions to measuring and conceptualizing BDAC and its performance implications. Constant performers, including Gupta and George (2016), Wamba et al. (2017), and Sivarajah et al. (2017), maintain enduring influence over time. Their contributions include hierarchical models of BDAC, empirical validation of performance outcomes, and syntheses of adoption challenges.

Gupta and George (2016) developed the widely adopted formative measurement model of BDAC as a third-order construct comprising tangibles (basic resources, data, technology), human skills (technical and managerial expertise), and intangibles (organizational learning, data-driven culture). Wamba et al., (2017) empirically validated the hierarchical BDAC model, highlighting both direct and indirect effects on firm market and financial performance mediated by process-oriented dynamic capabilities. Mikalef and Pateli (2017) extended prior work by linking IT-enabled dynamic capabilities with firm agility and competitive performance, employing fuzzy-set qualitative comparative analysis (fsQCA) to delineate boundary conditions for BDAC's value creation. Sivarajah et al. (2017) conducted a systematic review that categorized organizational challenges in adopting BDA into data-related, process-related, and management-related areas, offering a synthesis of mitigation strategies.

Stylized Fact Three: BDAC is a multifaceted construct with significant potential to enhance organizational performance both directly and indirectly. However, successful adoption is contingent on overcoming challenges across data quality, process integration, and

managerial oversight

Constant Performers and Hot Papers

Two publications were identified as both constant performers and hot papers, indicating sustained citation relevance coupled with initial rapid uptake, a pattern indicative of influential and pioneering scholarship. Seddon and Currie (2017) and Wang and Hajli (2017) display dual characteristics, showing immediate influence with sustained relevance. Their works extend traditional BDAC models and validate industry-specific applications, exemplifying pioneering scholarship that bridges conceptual innovation and empirical testing.

Seddon and Currie (2017) extended the traditional 3Vs model of big data (volume, velocity, variety) to a 7Vs framework, applying it to high-frequency trading (HFT) firms and demonstrating how these organizations leverage BDAC for competitive advantage relative to low-frequency trading (LFT) counterparts. Wang & Hajli, (2017) applied resource-based theory (RBT) to develop a big data analytics-enabled business value model specific to healthcare, empirically validating the causal pathways through which BDAC generates business value across 63 healthcare organizations.

Stylized Fact Four: BDAC has materially enhanced competitive advantage across diverse industries, notably finance and healthcare, by strengthening firms' capabilities to link big data-derived insights with sustained value creation and strategic success.

The citation trajectories reveal several insights for BDAC theory development. Sleeping beauties highlight latent conceptual contributions that later become foundational, illustrating how theory often requires empirical validation or contextual adaptation before recognition. Hot papers reveal emergent empirical interest, signaling areas where BDAC frameworks are actively operationalized and debated. Constant performers show theoretical consolidation, where models and measures achieve enduring influence, reflecting consensus and validation in the field. Dual trajectory works demonstrate the integration of conceptual and empirical influence, underscoring the synergistic development of BDAC knowledge.

Table 1. Citation Trajectory Classification of BDAC Literature

Cited Reference	Trajectory Type
Gandomi and Haider (2015)	Sleeping beauty
Queiroz and Telles (2018)	Sleeping beauty
Markus (2015)	Hot paper
Xu, Frankwick, and Ramirez (2016)	Hot paper
Battistella, De Toni, De Zan, and Pessot (2017)	Hot paper
Mikalef, Pateli, and van de Wetering (2016)	Hot paper
Wang, Kung, and Byrd (2018)	Hot paper
Phillips-Wren, Iyer, Kulkarni, and Ariyachandra (2015)	Hot paper
Mikalef and Pateli (2016)	Hot paper
Wamba et al. (2017)	Constant performer
Mikalef and Pateli (2017)	Constant performer
Gupta and George (2016)	Constant performer
Sivarajah, Kamal, Irani, and Weerakkody (2017)	Constant performer
Seddon and Currie (2017)	Constant performer + Hot paper
Wang and Hajli (2017)	Constant performer + Hot paper

From a theoretical perspective, the CSA results suggest that BDAC research has evolved through a progressive interplay of conceptual proposals, empirical testing, and validation across contexts, aligning with resource-based and dynamic capability frameworks. Fragmentation remains in emerging subdomains, particularly cross-industry applications and socio-technical considerations.

Future Research Agenda: We propose actionable thematic tracks:

- **Measurement:** Refine BDAC operationalization and metrics across industries and contexts.
- **Mechanisms:** Investigate the processes through which BDAC translates into performance, including moderating and mediating factors.
- **Contextual Applications:** Explore cross-industry and cross-country variations in BDAC adoption and impact.
- **Ethics and Governance:** Examine negative externalities, privacy risks, and social implications of BDAC deployment.

Key Methodological Clarifications: Database choice (Scopus) may influence detected trajectories, with underrepresentation of regional or non-English sources. CSA provides descriptive insights into citation patterns rather than causal claims. Overlapping classifications reflect multi-dimensional influence of seminal works, based on CRExplorer output and manual confirmation.

This structured results and discussion revision links citation trajectories to conceptual and empirical maturation of BDAC, provides analytical interpretation for each trajectory, and translates bibliometric patterns into actionable theoretical and practical insights.

Contributions and an Agenda for Future Research

This study makes several significant contributions to the Big Data Analytics Capability (BDAC) literature and identifies critical gaps that warrant further scholarly inquiry. The discussion below details theoretical and practical implications supported by empirical evidence and outlines a structured agenda for future research.

Theoretical Contributions

This research offers the first systematic historical mapping of BDAC scholarship by utilizing Citation Sequence Analysis (CSA), which reveals the temporal citation trajectories of key contributions within the field. Previous bibliometric studies (Batistič & van der Laken, 2019b; Li & Ye, 2016) have explored broad big data research but lacked a focused historical analysis specific to BDAC. The identification of “sleeping beauties,” “hot papers,” and “constant performers” uncovers the evolutionary path of BDAC, from its conceptual infancy to its current multifaceted theoretical maturity. For example, seminal works by Gupta and George (2016) have sustained long-term influence (constant performers), whereas studies like Gandomi and Haider (2015) experienced delayed recognition but have since shaped foundational understanding of big data attributes.

The results confirm that BDAC is not a monolithic construct but rather a higher-order dynamic capability comprising tangible resources (e.g., data infrastructure), human skills (technical and managerial), and intangible assets (organizational learning, data-driven culture) (Gupta & George, 2016). This aligns with dynamic capability theory (Teece et al., 1997), which emphasizes an organization’s ability to integrate, build, and reconfigure resources in rapidly changing environments. The empirical validations by Wamba et al. (2017) and Mikalef and Pateli

(2017) further demonstrate BDAC’s positive influence on market performance and operational agility, supporting the view that BDAC underpins firm competitiveness through strategic resource orchestration.

While most research underscores BDAC’s benefits, this study highlights its “dark side,” including disruptive labor market effects, privacy concerns, and ethical dilemmas (Choi & Park, 2022; Markus, 2015). The inclusion of Markus’s (2015) findings about automation-induced job losses and privacy vulnerabilities emphasizes that BDAC adoption may generate significant externalities affecting employees and communities. These insights call for theoretical models that integrate socio-technical systems perspectives (Baxter & Sommerville, 2011) and critical information systems research (CISR) approaches to balance technological innovation with social responsibility.

Practical Implications

This study offers actionable insights for practitioners navigating BDAC adoption and deployment. The nuanced understanding of BDAC as a multidimensional capability stresses the importance of investing not only in technology but also in developing skilled personnel and cultivating a data-driven organizational culture (L. Chen et al., 2022). Moreover, the recognition of negative consequences alerts managers to proactively address challenges such as workforce reskilling, data governance, and privacy compliance to ensure sustainable BDAC implementation. Additionally, sector-specific findings, such as Wang et al.’s (2018) healthcare analytics framework and Seddon and Currie’s (2017) financial trading models, provide industry-relevant blueprints for leveraging BDAC to enhance competitive advantage.

Agenda for Future Research

Exploring the Capability Development Process and Pathways: Though foundational studies emphasize BDAC’s importance, few have explicated the processes through which firms develop this capability. Future research should employ longitudinal and qualitative methodologies to uncover mechanisms of resource accumulation, integration, and learning over time (Eisenhardt & Martin, 2000). Questions such as how firms overcome path dependencies or leverage prior IT capabilities to develop BDAC deserve particular attention (Mikalef et al., 2021).

Investigating the Negative Externalities and Stakeholder Impact: While positive outcomes dominate extant research, systematic empirical studies examining BDAC’s adverse impacts remain sparse. Future work should assess how BDAC contributes to labor displacement (Frey & Osborne, 2017), exacerbates privacy risks (Acquisti et al., 2016), or induces technostress among employees (Ragu-Nathan et al., 2008). The development of frameworks for ethical BDAC governance and stakeholder-inclusive impact assessments can guide responsible practice.

Understanding Adoption Barriers and Overcoming Implementation Challenges: Empirical evidence points to numerous challenges in BDAC adoption, including organizational inertia, data quality issues, and skill shortages (Queiroz & Telles, 2018; Sivarajah et al., 2017). Future studies should focus on identifying critical success factors and change management strategies, potentially drawing on innovation diffusion theories (Rogers, 2003) and technology acceptance models (Venkatesh et al., 2003) to explain heterogeneity in adoption outcomes.

Conducting Cross-Industry Comparative Studies to Build Grand Theories: The fragmented nature of BDAC research across industries calls for comparative studies that examine how sector-specific contingencies influence BDAC development and impact. For instance, financial services

(Seddon & Currie, 2017) and healthcare (Wang et al., 2018) have established BDAC's value, but less is known about its role in education, hospitality, or retail sectors. Comparative, multi-industry investigations can inform the creation of more integrative, generalizable theoretical frameworks explaining BDAC's role in diverse industrial ecosystems.

Conclusion

This study acknowledges several limitations that may shape the interpretation and generalizability of its findings. First, the bibliometric analysis relied exclusively on the Scopus database, which, despite its extensive journal coverage and rigorous indexing, may underrepresent non-English publications or regional journals. This limitation suggests that some influential BDAC contributions could be omitted, potentially biasing trajectory detection. Future research should consider triangulating multiple databases, such as Web of Science and Google Scholar, to provide a more comprehensive and externally valid representation of BDAC scholarship.

Second, the study utilized CRExplorer to perform Citation Sequence Analysis (CSA). While CRExplorer enables robust disambiguation and trajectory classification, alternative bibliometric platforms, including CitNetExplorer, offer complementary strengths, such as visualizing citation network interconnections and mapping co-citation clusters. Employing multiple analytical tools in future studies could enrich understanding of both structural and temporal patterns in BDAC research.

Third, the methodological focus was limited to CSA and the associated Reference Publication Year Spectroscopy (RPYS). Although these approaches provide insights into citation trajectories and historical influence, they primarily offer descriptive rather than causal inferences. Future investigations could integrate complementary bibliometric methods, including co-citation analysis or bibliographic coupling, to reveal relational structures among BDAC publications and refine theoretical understanding of knowledge accumulation.

Fourth, the analysis considered BDAC research at a general level, without delving into domain-specific subfields such as marketing analytics, supply chain

analytics, or predictive analytics. These subdomains may exhibit distinct developmental trajectories and contextualized patterns of influence. Conducting trajectory analyses within subfields could uncover nuanced trends and inform theory-building tailored to specific organizational applications.

In conclusion, BDAC represents a dynamic, multidimensional, and rapidly evolving area of scholarly inquiry, where foundational conceptual contributions, emergent empirical studies, and sustained influence collectively shape its development. By applying a robust, methodologically rigorous CSA approach, this study identifies and synthesizes key publications, traces historical and intellectual patterns, and translates citation trajectories into actionable insights for theory and practice. Despite its limitations, the study provides a structured research agenda to guide future investigation across measurement, mechanisms, contextual applications, and ethical considerations. Advancing this agenda will deepen understanding of BDAC and support responsible, evidence-informed application in organizational contexts, fulfilling the scholarly imperative to translate knowledge into action.

Acknowledgements

The authors acknowledge the use of artificial intelligence (AI) technologies, specifically ChatGPT (OpenAI) and Consensus, in the development of this manuscript. These tools were employed to assist with generating summaries of relevant literature and datasets, refining the articulation of practical applications and recommendations derived from the results, and enhancing the clarity and readability of the manuscript through improvements to grammar, structure, and overall language quality. AI was also used beyond standard proofreading functions to support substantive revisions to the content and organization of the text. Despite the use of AI tools in these capacities, the authors assert that all intellectual contributions, analytical reasoning, and interpretations presented in this article are their own. The authors remain solely responsible for the originality, accuracy, and integrity of the content, in full accordance with the ethical standards outlined by the Committee on Publication Ethics (COPE).

References

- Acquisti, A., Taylor, C., & Wagman, L. (2016). The Economics of Privacy. *Journal of Economic Literature*, 54(2), 442–492. <https://doi.org/10.1257/jel.54.2.442>
- Argote, L., & Miron-Spektor, E. (2011). Organizational Learning: From Experience to Knowledge. *Organization Science*, 22(5), 1123–1137. <https://doi.org/10.1287/orsc.1100.0621>
- Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a Curated, High-Quality Bibliometric Data Source. *Quantitative Science Studies*, 1(1), 377–386. https://doi.org/10.1162/qss_a_00019
- Batistić, S., & van der Laken, P. (2019a). History, Evolution and Future of Big Data and Analytics. *British Journal of Management*, 30(2), 229–251. <https://doi.org/10.1111/1467-8551.12339>
- Batistić, S., & van der Laken, P. (2019b). History, Evolution and Future of Big Data and Analytics. *British Journal of Management*, 30(2), 229–251. <https://doi.org/10.1111/1467-8551.12339>
- Battistella, C., De Toni, A. F., De Zan, G., & Pessot, E. (2017). Cultivating Business Model Agility Through Focused Capabilities. *Journal of Business Research*, 73, 65–82. <https://doi.org/10.1016/j.jbusres.2016.12.012>
- Baxter, G., & Sommerville, I. (2011). Socio-Technical Systems: From Design Methods to Systems Engineering. *Interacting with Computers*, 23(1), 4–17. <https://doi.org/10.1016/j.intcom.2010.07.003>
- Brinch, M., Gunasekaran, A., & Wamba, S. F. (2021). Firm-Level Capabilities Towards Big Data Value Creation. *Journal of Business Research*, 131, 539–548. <https://doi.org/10.1016/j.jbusres.2021.03.059>
- Carr, N. G. (2003). IT Doesn't Matter. *Educause Review*, 38(6), 24–38.
- Cetindamar, D., Shdifat, B., & Erfani, S. (2020). Assessing Big Data Analytics Capability and Sustainability in Supply Chains. *Proceedings of the 53rd Hawaii International Conference on System Sciences*.
- Chen, C. (2003). *Mapping Scientific Frontiers: The Quest for Knowledge Visualization*. Springer.
- Chen, L., Liu, H., Zhou, Z., Chen, M., & Chen, Y. (2022). IT-business alignment, big data analytics capability, and strategic decision-making: Moderating roles of event criticality and disruption of COVID-19. *Information & Management*. <https://doi.org/10.1016/j.im.2022.103745>
- Choi, H.-Y., & Park, J. (2022). Do data-driven CSR initiatives improve CSR performance? The importance of big data analytics capability. *Technological Forecasting and Social Change*, 182, 121802. <https://doi.org/10.1016/j.techfore.2022.121802>
- Ciasullo, M. V., Montera, R., & Douglas, A. (2022). Building SMEs' resilience in times of uncertainty: The role of big data analytics capability and co-innovation. *Transforming Government: People, Process and Policy*, 16(2), 203–217. <https://doi.org/10.1108/TG-07-2021-0120>
- Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science*

- Quarterly*, 35(1), 128–152. <https://doi.org/10.2307/2393553>
- Davenport, T. H. (2006). Competing on Analytics. *Harvard Business Review*, 84(1), 98–107.
- Davenport, T. H., & Harris, J. G. (2017). *Competing on Analytics: The New Science of Winning*. Harvard Business Press.
- De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is Big Data? A Consensual Definition. *Journal of Business Research*, 70, 263–286. <https://doi.org/10.1016/j.jbusres.2016.08.001>
- Donthu, N., Kumar, S., Pattnaik, D., & Lim, W. M. (2021). How to Conduct a Bibliometric Analysis. *Journal of Business Research*, 133, 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic Capabilities: What Are They? *Strategic Management Journal*, 21(10–11), 1105–1121.
- Emmanuel, I., & Stanier, C. (2016). Defining Big Data. *Proceedings of the International Conference on Big Data and Advanced Wireless Technologies*, 1–6.
- Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data Consumer Analytics and the Transformation of Marketing. *Journal of Business Research*, 69(2), 897–904. <https://doi.org/10.1016/j.jbusres.2015.07.001>
- Fang, B., & Zhang, P. (2016). Big Data in Finance. In *Big Data Concepts, Theories, and Applications* (pp. 391–412). Springer. https://doi.org/10.1007/978-3-319-27763-9_11
- Frey, C. B., & Osborne, M. A. (2017). The Future of Employment. *Technological Forecasting and Social Change*, 114, 254–280. <https://doi.org/10.1016/j.techfore.2016.08.019>
- Gandomi, A., & Haider, M. (2015). Beyond the Hype: Big Data Concepts and Analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- George, G., Haas, M. R., & Pentland, A. (2014). Big Data and Management. *Academy of Management Journal*, 57(2), 321–326. <https://doi.org/10.5465/amj.2014.4002>
- Gupta, M., & George, J. F. (2016). Toward the Development of a Big Data Analytics Capability. *Information & Management*, 53(8), 1049–1064. <https://doi.org/10.1016/j.im.2016.07.004>
- Hao, S., Zhang, H., & Song, M. (2019). Big data, big data analytics capability, and sustainable innovation performance. *Sustainability*, 11(24), 7145. <https://doi.org/10.3390/su11247145>
- Haunschild, R., & Borrmann, L. (2022). Reference publication year spectroscopy (RPYS) in practice: A software tutorial. *Scientometrics*. <https://doi.org/10.1007/s11192-022-04369-8>
- KPMG. (2022). *So You Have Alteryx, What's Next? The Hard Truth About Spreadsheets and Data Transformation in the 21st Century*. <https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2022/so-you-have-alteryx-what-next.pdf>
- Li, J., & Ye, F. Y. (2016). Distinguishing sleeping beauties in science. *Scientometrics*, 108(2), 821–828. <https://doi.org/10.1007/s11192-016-1883-7>
- Maltby, D. (2011). Big Data Analytics. *Proceedings of the 74th ASIS&T Annual Meeting*, 48(1).
- Markus, M. L. (2015). New games, new rules, new scoreboards: The potential consequences of big data. *Journal of Information Technology*, 30(1), 58–59. <https://doi.org/10.1057/jit.2014.28>
- Mikalef, P., van de Wetering, R., & Krogstie, J. (2021). Building Dynamic Capabilities by Leveraging Big Data Analytics. *Information & Management*, 58(6). <https://doi.org/10.1016/j.im.2021.103412>
- Naeem, M., Jamal, T., Diaz-Martinez, J., Butt, S. A., Montesano, N., Tariq, M. I., & De-La-Hoz-Valdiris, E. (2022). Trends and Future Perspective Challenges in Big Data. In *Advances in Intelligent Data Analysis and Applications* (pp. 309–325). Springer. https://doi.org/10.1007/978-981-16-7862-0_23
- Nguyen, T., Li, Z., Spiegler, V., Ieromonachou, P., & Lin, Y. (2018). Big data analytics in supply chain management: A state-of-the-art literature review. *Computers & Operations Research*, 98, 254–264. <https://doi.org/10.1016/j.cor.2017.07.004>
- Phillips-Wren, G., Iyer, L. S., Kulkarni, U., & Ariyachandra, T. (2015). Business analytics in the context of big data: A roadmap for research. *Communications of the Association for Information Systems*, 37(1), 23. <https://doi.org/10.17705/1CAIS.03723>
- Queiroz, M. M., & Telles, R. (2018). Big data analytics in supply chain and logistics: An empirical approach. *The International Journal of Logistics Management*, 29(2), 767–783. <https://doi.org/10.1108/IJLM-05-2017-0114>
- Ragu-Nathan, T. S., Tarafdar, M., Ragu-Nathan, B. S., & Tu, Q. (2008). The consequences of technostress for end users in organizations: Conceptual development and empirical validation. *Information Systems Research*, 19(4), 417–433. <https://doi.org/10.1287/isre.1070.0165>
- Russom, P. (2011). *Big Data Analytics: TDWI Best Practices Report, Fourth Quarter 2011*. <https://tdwi.org/tdwi/research/bpr/2011>
- Sahoo, S. (2021). Big data analytics in manufacturing: A bibliometric analysis of research in the field of business management. *International Journal of Production Research*, 59(15), 4561–4585. <https://doi.org/10.1080/00207543.2020.1865580>
- Santarsiero, F., Jarrar, Y., & Carlucci, D. (2021). Creating value from Big Data: a knowledge assets-based view. *Knowledge Management Research & Practice*, 21, 714–724. <https://doi.org/10.1080/14778238.2021.2015264>
- Seddon, J. J. M., & Currie, W. L. (2017). A model for unpacking big data analytics in high-frequency trading. *Journal of Business Research*, 70, 300–307. <https://doi.org/10.1016/j.jbusres.2016.08.022>
- Shi, Y. (2022). *Big Data Analytics: Techniques and Applications*. Springer. <https://doi.org/10.1007/978-3-030-12345-6>
- Singh, S., & El-Kassar, A.-N. (2019). Role of big data analytics in developing sustainable capabilities. *Journal of Cleaner Production*, 213, 1264–1273. <https://doi.org/10.1016/j.jclepro.2018.12.199>
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research*, 70, 263–286. <https://doi.org/10.1016/j.jbusres.2016.08.001>
- Tan, K. H. (2018). Relationship of big data analytics capability and product innovation performance using SmartPLS 3.2.6: Hierarchical component modelling in PLS-SEM. *International Journal of Supply Chain Management*, 7(1), 1838. <https://doi.org/10.59160/ijscm.v7i1.1838>
- Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic Capabilities and Strategic Management. *Strategic Management Journal*, 18(7), 509–533. [https://doi.org/10.1002/\(SICI\)1097-0266\(199708\)18:7<509::AID-SMJ882>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z)
- Thor, A., Borrmann, L., Haunschild, R., Leydesdorff, L., Mutz, R., & Marx, W. (2018). *CitedReferencesExplorer (CRExplorer) Manual (Version 1.9)*. <https://andreas-thor.github.io/cre/manual.pdf>
- Thor, A., Borrmann, L., Marx, W., & Mutz, R. (2018). Identifying single influential publications in a research field: New analysis opportunities of the CRExplorer. *Scientometrics*, 116(1), 591–608. <https://doi.org/10.1007/s11192-018-2733-7>
- Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the longwall method of coal-getting. *Human Relations*, 4(1), 3–38. <https://doi.org/10.1177/001872675100400101>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Waheed, A., Yang, Q., & Khan, S. (2018). Investment in intangible resources and capabilities spurs sustainable competitive advantage and firm performance. *Corporate Social Responsibility and Environmental Management*. <https://doi.org/10.1002/csr.1678>
- Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-F., Dubey, R., & Childe, S. J. (2017). Big Data Analytics and Firm Performance. *Journal of Business Research*, 70, 356–365. <https://doi.org/10.1016/j.jbusres.2016.08.009>
- Wang, Y., & Hajli, N. (2017). Exploring the path to big data analytics success in healthcare. *Journal of Business Research*, 70, 287–299. <https://doi.org/10.1016/j.jbusres.2016.08.002>
- Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3–13. <https://doi.org/10.1016/j.techfore.2015.12.019>
- Xu, Z., Frankwick, G. L., & Ramirez, E. (2016). Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of Business Research*, 69(5), 1562–1566. <https://doi.org/10.1016/j.jbusres.2015.10.017>
- Zehir, C., Karaboğa, T., & Başar, D. (2020). The Transformation of Human Resource Management and Its Impact on Overall Business Performance: Big Data Analytics and AI Technologies in Strategic HRM. In *Digital Business Strategies in Blockchain Ecosystems: Transformational Design and Future of Global Business* (pp. 265–279). Springer.