

Regulatory Catalysis and Scholarly Shift: A Bibliometric Analysis of Blockchain Emergence in Korea's Post-Reform Data Privacy Landscape

Yong-Jae Lee*

ARTICLE INFO

Article history:
Online First 11 September 2026

Keywords:
Data Privacy Reform,
Blockchain Innovation,
Bibliometric Analysis,
Regulatory Catalysis,
Topic Modeling (STM),
Decentralized Governance

ABSTRACT

This study empirically investigates how South Korea's 2020 "Data Bills" revision functioned as a regulatory catalyst reshaping academic research priorities, employing a longitudinal bibliometric analysis of 455 Korean papers (1998–2022) through four integrated methods: Bai-Perron structural break testing, TF-IDF vocabulary analysis, weighted log-odds ratio analysis, and Structural Topic Modeling (STM, K=8) combined with keyword co-occurrence Social Network Analysis. Post-reform research output increased 41% by 3-year moving average, with blockchain emerging as the strongest post-reform lexical signal (log-odds: 0.568) and the fastest-growing research topic with a 53% rise in prevalence, while civic (0.553) and service (0.553) signals indicated a decisive shift from technocratic security discourse toward citizen-facing governance frameworks, and AI adoption remained entirely undetectable throughout the corpus. These findings offer policymakers a replicable, evidence-based bibliometric framework for evaluating the downstream scholarly effects of digital governance reforms. As the analysis is based on publication patterns rather than direct measures of societal or industrial outcomes, these results should be interpreted as evidence of shifting research priorities rather than confirmed societal impact. Korea's regulatory experience nonetheless offers a transferable case study for jurisdictions undertaking similar data regulation transitions, though the relatively small post-reform corpus (n=82) warrants further longitudinal validation.

1. Introduction

1.1 Research Background

The exponential growth of the data industry has positioned personal data as a critical asset for economic innovation and digital transformation. In South Korea, this rapid expansion necessitated a robust legal framework for personal information protection, leading to the enactment of the Personal

* Postdoctoral Researcher, Department of Industrial Data Engineering, Hanyang University (yj11021@hanyang.ac.kr)(First & Corresponding Author)



Information Protection Act (PIPA) in 2011 (Chik, 2013; Greenleaf & Park, 2014; Ko & Park, 2020). This Act established a comprehensive regulatory system for data collection, use, and disclosure (Oh et al., 2020; Zoo et al., 2013). However, over the past decade, industry's evolution has created a significant tension: the drive for data-driven innovation increasingly demanded more flexible use of personal data, while existing regulations were perceived as restrictive. This conflict highlighted an urgent need for regulatory reform to balance the competing imperatives of privacy protection and industrial growth. Against this backdrop, the Korean National Assembly undertook a significant legislative overhaul in 2020, revising the so-called "Data Bills"—comprising the Personal Information Protection Act, the Information and Communication Network Act, and the Credit Information Act (Lee & Jung, 2021; Lee & Jung, 2020). This revision aimed to reorganize the regulatory landscape, fostering a more conducive environment for the data industry while maintaining core privacy principles. This legislative shift represents a pivotal moment, triggering significant changes in the technology market related and scholarly discourse. Therefore, a systematic analysis of the societal impacts of these changes, both before and after the revision, is crucial to understanding the consequences of this regulatory evolution.

1.2 Research Objectives and Questions

The primary aim of this study is to systematically examine, through bibliometric evidence, how the transformation within the Korean data industry's technology market following the revision of the "Data Bills" is reflected in scholarly research and discourse. To achieve this overarching goal, the research establishes three specific objectives. First, it seeks to systematically investigate the tangible effects of the regulatory changes on the data technology market and its broader social implications. Second, the study aims to identify and analyze the evolution of academic research trends in personal information protection before and after the legislative revision, utilizing bibliographic analysis and Social Network Analysis (SNA) of keyword patterns.

Third, it endeavors to investigate the role of technological innovation, specifically the adoption of blockchain security systems, in enhancing data protection and fostering community trust in the post-reform environment. By synthesizing findings from these objectives, this research will provide a comprehensive understanding of the dynamic interplay between regulation, technology, and society in the contemporary Korean data landscape. Based on the stated objectives, this study is guided by the following research questions.

RQ1: How are the changes in the Korean data industry-related technology market, following the revision of the Data Bills, reflected in the scholarly research landscape?

RQ2: How have research trends in personal information protection, as revealed through bibliographic and SNA keyword analysis, shifted after the regulatory reform?

RQ3: To what extent does the implementation of blockchain-based security solutions after the Data Bills revision contribute to creating a more secure and socially active digital community?

The significance of this study lies in its timely and multifaceted investigation of a major regulatory transition in a leading digital economy. By employing bibliographic and Social Network Analysis (SNA) of research papers, it offers a novel, data-driven perspective on how scholarly priorities evolve in response to policy changes. This methodology provides quantifiable evidence of shifting research paradigms, moving beyond theoretical discussion. Furthermore, the study critically examines the societal implications of these changes, focusing on the real-world integration of technologies like blockchain to address the shortcomings of centralized security systems. The findings are expected to provide valuable insights for policymakers in assessing the impact of the Data Bills and guiding future regulatory fine-tuning. For industry stakeholders, it highlights emerging technological trends and societal expectations. Ultimately, this research contributes to the global discourse on balancing data innovation with personal information protection, offering a case study of Korea's attempt to navigate this complex terrain.

2. Literature Review

This chapter synthesizes the existing literature to establish the foundational context for this study. It begins by examining the growth and regulatory landscape of the Korean data industry, which sets the stage for the subsequent legislative reforms. It then details the evolution of Korea's personal information protection framework, culminating in the revision of the Data Bills. Finally, it explores the intersection of technology and society, reviewing the potential of blockchain and the shifting research trends in response to these regulatory changes.

2.1 The Korean Data Industry and Its Regulatory Evolution

South Korea's data industry has witnessed remarkable growth, becoming a cornerstone of the nation's economy driven by rapid technological advancements and digital transformation (Alsharif et al., 2020; López-Muñoz et al., 2013). The industry encompasses a diverse range of sectors including data collection, storage, processing, analysis, and utilization, and it plays a critical role in fueling innovation across various domains such as finance, healthcare, manufacturing, and retail (Li et al., 2020; Shin, 2016; Yang & Gu, 2021). The market value reached an estimated 23.1 trillion South Korean won in 2021, indicating a significant growth trajectory fueled by the demand for data-driven solutions and the adoption of AI and ML technologies (Manyika et al., 2011; Rübmann et al., 2015). This data-driven paradigm enhances productivity, optimizes operations, and enables the development of new products and services, with applications ranging from fraud detection in finance to personalized medicine in healthcare (Abdou et al., 2016; Liuska, 2021; Orlova, 2021; Yap et al., 2011).

Parallel to the industry's growth, the Korean government established a comprehensive regulatory framework to protect personal information. The primary law is the Personal Information Protection Act (PIPA), enacted in 2011, which sets forth principles like purpose limitation, data minimization,

and transparency for the collection, use, and disclosure of personal information (Greenleaf & Park, 2014; Ko et al., 2017; Ko & Park, 2020). It mandates obtaining consent and implementing appropriate security measures. This framework is supplemented by the Information and Communication Network Act (ICNA), which includes provisions against unauthorized data collection, and the Credit Information Act (CIA), which specifically regulates credit information with principles of accuracy and purpose limitation (Kuczerawy, 2019; Rhie, 2013; Shinwan & Kim, 2018). Alongside these laws, government-issued guidelines and institutions like the Personal Information Protection Commission (PIPC) provide detailed compliance guidance and enforcement, making Korea a leading country in this field (Joo & Kwon, 2023).

Recognizing the need to foster a data-driven economy while ensuring protection, the government enacted the landmark Framework Act on Promotion of Industry and Utilization of the Data (Data Framework Act) in 2021 (Baek & Park, 2015; Oh & Larson, 2011; Shin, 2016; Sung, 2018). This act aimed to establish a comprehensive regulatory framework for data collection, utilization, and protection, promoting data infrastructure and innovation (Cavanillas et al., 2016; Fast et al., 2023; van Ooijen et al., 2019). This was part of a broader legislative overhaul. The most significant revision of the "Data Bills"—comprising PIPA, ICNA, and CIA—took effect in 2020. This revision introduced key provisions such as strengthening requirements for obtaining consent, granting individuals the right to request data deletion and access, and imposing stricter security obligations on businesses. This revision aimed to resolve the tension between the restrictive nature of earlier regulations and the industry's demand for more flexible data use, marking a pivotal moment in Korea's regulatory landscape.

2.2 Blockchain Technology as a Catalyst for Secure Data Utilization

Blockchain technology has emerged as a transformative force with the potential to address data protection challenges exacerbated by the growth of the data industry (Li et al., 2023). Its decentralized, transparent, and tamper-proof nature offers a promising alternative to traditional centralized data management systems, which are often vulnerable to security breaches (Bansal et al., 2016). By employing a distributed ledger, blockchain ensures data is stored in an immutable manner, making it more resistant to unauthorized access or manipulation. This inherent security significantly bolsters trust in data handling practices, particularly for sensitive personal information, aligning with the enhanced security requirements of the revised Data Bills (Hasan et al., 2022).

Beyond security, blockchain technology facilitates greater accountability and traceability in data usage. Each transaction or data modification on the blockchain is recorded and visible to authorized participants, creating an auditable trail to track the movement and usage of personal data (Li et al., 2021). This transparency empowers individuals to monitor how their data is being used and holds organizations accountable for their data handling practices, directly supporting the principles of transparency and individual rights reinforced by the recent regulatory reforms (Henein et al., 2020). The revision of the Data Bills thus creates a significant opportunity to examine the role of blockchain in achieving the dual goals of safeguarding privacy and enabling responsible data usage. The potential benefits of blockchain extend beyond personal data protection, encompassing

a broader range of societal applications. Its decentralized and trustless nature can foster collaboration and innovation in various domains, from supply chain management to healthcare (Li et al., 2023). By enabling secure and transparent data sharing among trusted participants, blockchain can facilitate more efficient and effective processes, leading to improved outcomes across various sectors and contributing to a more secure and data-driven society.

2.3 Evolving Research Trends and Social Impact

Prior to the revision of the Data Bills, the lack of comprehensive regulations led to significant concerns regarding data breaches, misuse of personal data, and the erosion of individual privacy (Kim et al., 2017). While the enactment of PIPA in 2011 was a step forward, it was not entirely sufficient to address all concerns. This environment influenced research trends, which predominantly focused on the risks associated with data processing and the development of new technologies to protect personal information. There was relatively little research on the potential benefits of using personal data or the social implications of data-driven innovation (Bennett, 1992).

The 2020 revision of the Data Bills has had a profound impact on both the social landscape and academic research (Sung, 2018). The strengthened regulations have helped protect individual privacy, reduce the risk of data breaches, and increase transparency in the data industry. This regulatory shift is reflected in a corresponding evolution in research trends. There is now a growing body of research on the potential benefits of using personal data to develop new services and personalization experiences, indicating a more balanced view of data utilization (Setiawati et al., 2020). Furthermore, the revision has catalyzed an increase in research on blockchain technology as a means of protecting personal data, exploring its potential to create secure and tamper-proof systems in the new regulatory context (Henein et al., 2020). The current research landscape in personal information protection is thus characterized by a dual focus: investigating the social impact and benefits of data use within the new legal framework and exploring innovative technological solutions like blockchain to enforce privacy principles effectively. This shift demonstrates how policy changes can directly influence scholarly priorities, moving the discourse from purely defensive postures to more nuanced explorations of how to enable innovation responsibly. These changes are likely to have a continuing impact on the social role of technology in the Korean data industry for years to come.

Taken together, these three strands of literature reveal a gap that motivates the present study. Research on Korea's data industry and its regulatory evolution (Section 2.1) documents the structural tension between innovation and protection that the Data Bills revision sought to resolve, but offers little empirical evidence of how this legislative shift has actually reshaped the research agenda itself. Research on blockchain as a privacy-enhancing technology (Section 2.2) identifies its theoretical potential to address the shortcomings of centralized data systems, yet does not examine whether this potential has translated into a measurable shift in scholarly attention following regulatory change. Finally, prior work on evolving research trends (Section 2.3) establishes that policy shifts can influence scholarly priorities in principle, but has not applied a systematic, longitudinal bibliometric method to the Korean case specifically. This study addresses this gap by empirically tracing how the 2020 Data Bills revision is reflected in the scholarly literature itself—connecting the regulatory context

(2.1), the specific technology of interest (2.2), and the broader question of policy-driven research change (2.3) into a single, data-driven investigation.

3. Methodology

3.1 Research Design and Data Collection

This study employs a mixed-methods approach combining bibliometric analysis, social network analysis (SNA), and temporal trend analysis to examine the transformation of Korea's data industry following the 2020 Data Bills revision (Elbashbishy et al., 2022; Scott, 1988). The research design follows a longitudinal comparative framework, dividing the analysis period into two distinct phases: pre-reform (1998-2018) and post-reform (2019-2022). The corpus comprises 455 academic papers related to personal information protection, data governance, and technology policy in Korea. Papers were collected from KCI DB, covering publications indexed between 1998 and 2022. Search queries combined Korean and English keyword terms—including "personal information protection," "data privacy," "PIPA," and "블록체인/blockchain"—applied to title, abstract, and keyword fields. Papers were included if they addressed personal data protection, data governance, or related technology policy in the Korean context; papers with no substantive engagement with these themes were excluded after title/abstract screening. Retrieved papers were then classified by publication year to enable temporal comparison across the pre-reform (1998–2018) and post-reform (2019–2022) periods. This 25-year span provides sufficient temporal depth to capture both baseline trends and post-reform changes.

3.2 Text Processing and Preprocessing

3.2.1 Tokenization and Normalization

The text preprocessing pipeline follows established natural language processing protocols (Kostoff et al., 2001; Nowakowska, 2025). Raw text data underwent tokenization using the tidytext package in R, followed by lemmatization using the textstem library to reduce words to their root forms. Stop words were removed based on standard English stop word lists, and terms with fewer than three characters were excluded to eliminate noise. The preprocessing function can be formally expressed in Equation 1. Where T represents the original text, T' is the processed token set, w represents individual words, and S is the stop word set.

$$T' = \text{lemmatize}(\text{filter}(\text{tokenize}(T), |w| \geq 3 \wedge w \notin S)) \quad (1)$$

3.2.2 Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF analysis was employed to identify terms that are both frequent within specific time periods and distinctive across periods. The TF-IDF score for term t in document d within collection D

is calculated in Equations 2 and 3. Here, $f(t,d)$ represents the frequency of term t in document d , $|d|$ is the total number of terms in document d , $|D|$ is the total number of documents, and $||[d \in D: t \in d]||$ is the number of documents containing term t .

$$TF-IDF(t,d,D) = TF(t,d) \times IDF(t,D) \quad (2)$$

$$TF(t,d) = \frac{f(t,d)}{|d|}, \quad IDF(t,D) = \log \frac{|D|}{|[d \in D: t \in d]|} \quad (3)$$

3.2.3 Log-Odds Ratio Analysis

To identify terms that show significant differential usage between periods, we employed weighted log-odds ratios using the tidylo package. Log-odds ratio for term t comparing period A to period B is Equation 4. where $n_A(t)$ and $n_B(t)$ are the frequencies of term t in periods A and B respectively, N_A and N_B are the total word counts in each period, and α is a smoothing parameter (typically 0.5).

$$\log-odds(t) = \log \frac{n_A(t) + \alpha}{N_A - n_A(t) + \alpha} - \log \frac{n_B(t) + \alpha}{N_B - n_B(t) + \alpha} \quad (4)$$

3.3 Temporal Trend Analysis

To identify significant shifts in publication patterns, we applied the Bai-Perron structural break test using the strucchange package (Weideman et al., 2017). The test identifies multiple breakpoints in the time series by minimizing the sum of squared residuals in Equation 5. Where $T_1, T_2, \dots, T_m$ represent the breakpoints, y_t is the dependent variable at time t , x_t are the regressors, and β_i are the regime-specific parameters. To reduce noise in annual publication counts, we applied a 3-year moving average in Equation 6.

$$RSS(T_1, T_2, \dots, T_m) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} (y_t - x_t' \beta_i)^2 \quad (5)$$

$$MA_3(t) = \frac{y_{t-1} + y_t + y_{t+1}}{3} \quad (6)$$

3.4 Topic Modeling, Network Analysis, and Technology Adoption Metrics

3.4.1 Topic Modeling

We employed STM to identify latent topics and their evolution over time (Kuhn, 2018; Roberts et al., 2019). The STM incorporates document-level covariates to model topic prevalence as a function of time period and year in Equation 7. where θ_d represents the topic proportions for document d , X_d contains covariates (period and year), γ represents covariate effects, and Σ is the covariance matrix. The model was fitted with $K=8$ topics, incorporating both periods (before/after 2019) and

a smooth function of year as prevalence covariates in Equation 8. Topic labels were assigned based on manual review of the highest-probability terms (FREX and highest-probability word lists) and representative documents for each topic. We acknowledge that formal comparison of model fit across alternative values of K (e.g., using held-out likelihood or semantic coherence metrics) was not conducted in this study; this represents a methodological limitation, and sensitivity of the results to the choice of K is discussed further in Section 6.

$$\theta_d \sim \text{LogisticNormal}(\mu = X_d\gamma, \Sigma) \quad (7)$$

$$\text{prevalence} \sim \text{period} + s(\text{year}) \quad (8)$$

3.4.2 Social Network Analysis

Keyword co-occurrence networks were constructed using pairwise counting within documents. Two terms are connected if they co-occur in the same document, with edge weights representing co-occurrence frequency in Equation 9. Where $w(i, j)$ is the weight of the edge between terms i and j .

$$w(i, j) = |\{d : t_i \in d \wedge t_j \in d\}| \quad (9)$$

Multiple centrality measures were calculated to assess keyword importance in Equations 10 and 13. where A is the adjacency matrix, σ_{st} is the number of shortest paths between nodes s and t , $d(v, u)$ is the number of those paths passing through v , $d(v, u)$ is the shortest path distance, and λ is the largest eigenvalue of A .

$$\text{Degree Centrality} = C_D(v) = \sum_{u \in V} A(v, u) \quad (10)$$

$$\text{Betweenness Centrality} = C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (11)$$

$$\text{Closeness Centrality} = C_C(v) = \frac{1}{\sum_{u \in V} d(v, u)} \quad (12)$$

$$\text{Eigenvector Centrality} = x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} A_{v,t} x_t \quad (13)$$

3.4.3 Technology Adoption Metrics

To quantify the adoption of emerging technologies, we calculated technology-specific adoption scores in Equations 14 and 15. Where tech_terms represent the predefined vocabulary for each technology (blockchain, AI), and L represents document length.

$$Adoption \setminus - Score(tech, period) = \frac{\sum_{d \in period} Count(tech \setminus - terms, d)}{\sum_{d \in period} |d|} \quad (14)$$

$$Incidence \setminus - Rate(tech, period) = \frac{|[d \in period : Count(tech \setminus - terms, d) > 0]|}{|period|} \quad (15)$$

4. Results

4.1 Temporal Trends and Structural Changes

The analysis of publication trends reveals significant shifts in research focus following the 2020 Data Bills revision. The corpus of 455 papers spans from 1998 to 2022, with 373 papers published before 2019 and 82 papers after 2019. The structural break analysis identified multiple potential breakpoints, though the concentrated publication increases around 2020-2021 suggests a clear response to the legislative changes in Figure 1. The moving average analysis reveals a notable acceleration in publication activity following 2020, with the 3-year moving average increasing from 16.3 in 2019 to 23.0 in 2021, representing a 41% increase in research activity.

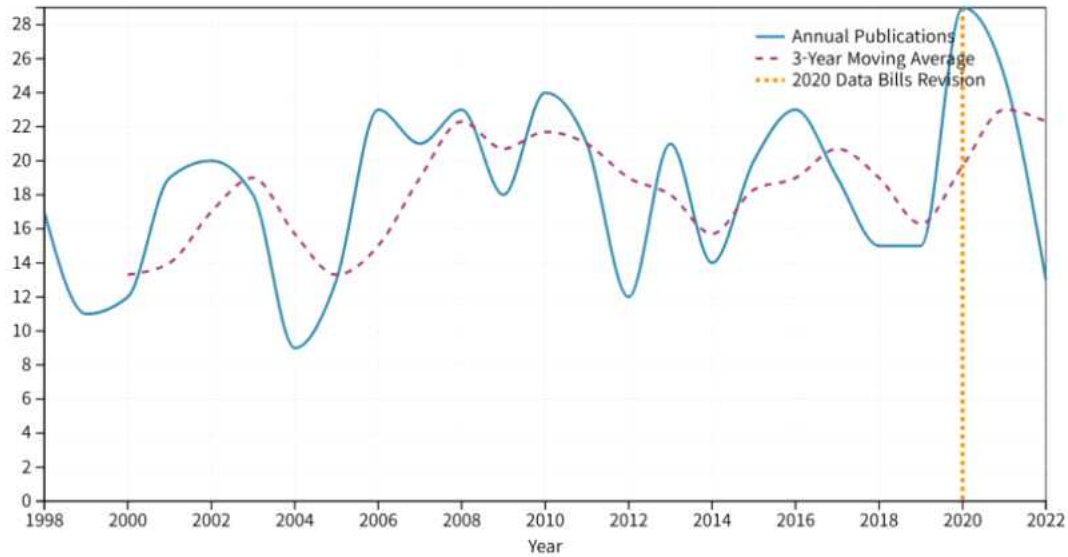


Fig. 1. Publication Trends and Moving Average (1998-2022)

4.2 Keyword Evolution and TF-IDF Analysis and Log-Odds Ratio Analysis

The TF-IDF analysis reveals substantial shifts in research vocabulary between the pre-reform and post-reform periods. Table 1 presents the top-ranked terms for each period, highlighting the emergence of new regulatory and technological discourse. The weighted log-odds analysis identifies

terms that show statistically significant differential usage between periods. The results (Table 2) reveal strong polarization in vocabulary, with traditional cybersecurity terms declining and block-chain-governance terms ascending.

Table 1. Top 15 Terms by TF-IDF Score

Rank	Before 2019	TF-IDF	After 2019	TF-IDF
1	privacy	0.0625	privacy	0.0625
2	data	0.0483	data	0.0483
3	security	0.0327	blockchain	0.0440
4	system	0.0284	study	0.0412
5	network	0.0270	security	0.0327
6	protection	0.0213	article	0.0284
7	information	0.0170	countermeasure	0.0284
8	technology	0.0170	externality	0.0284
9	policy	0.0127	fintech	0.0284
10	analysis	0.0127	map	0.0284
11	framework	0.0127	personal	0.0284
12	implementation	0.0127	civic	0.0270
13	development	0.0127	compliance	0.0270
14	research	0.0127	paper	0.0270
15	approach	0.0127	service	0.0270

Table 2. Top Differential Terms (Log-Odds Ratio)

Term	Period	Frequency	Log-Odds (Weighted)	Interpretation
blockchain	After 2019	31	0.568	Strong post-reform emergence
civic	After 2019	19	0.553	Increased civic engagement focus
service	After 2019	19	0.553	Platform service orientation
stack	After 2019	19	0.553	Technical architecture emphasis
compute	Before 2019	50	0.462	Traditional computing focus
edge	Before 2019	50	0.462	Infrastructure-centric vocabulary
asia	Before 2019	60	0.371	Regional comparative studies
regime	Before 2019	60	0.371	Regulatory regime analysis

4.3 Topic Modeling Results and Social Network Analysis

The STM analysis with K=8 topics reveals distinct thematic clusters and their evolution over time in Figure 2. Topic prevalence analysis shows significant shifts in research priorities following the legislative revision. The most significant finding is the emergence of "Blockchain Innovation" as a dominant topic, with a 53% increase in prevalence (from 0.147 to 0.225) following the Data Bills revision in Table 3.

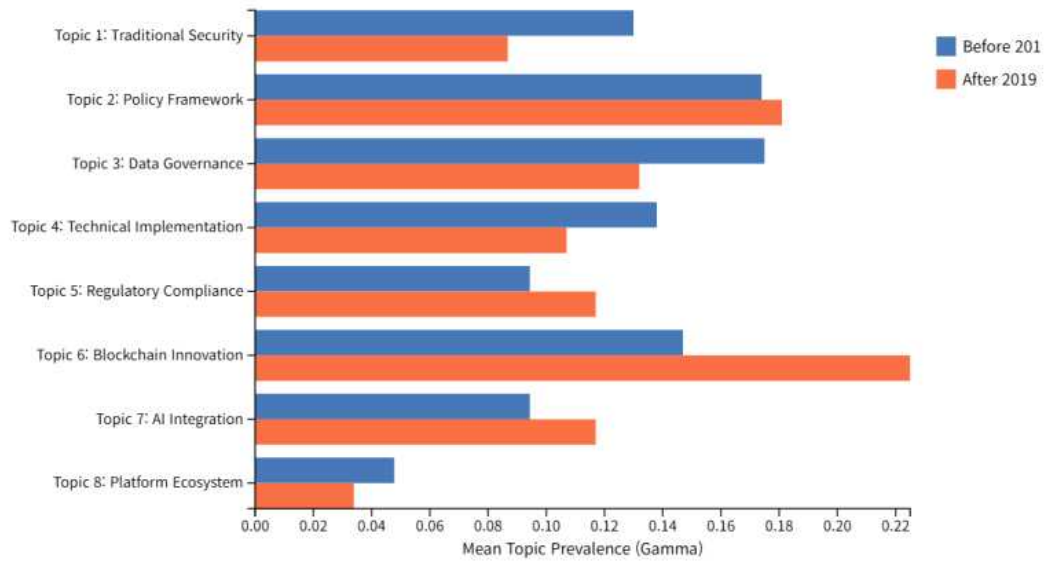


Fig. 2. Topic Prevalence by Period (Mean Gamma Values).

Table 3. Topic Evolution Summary

ID	Topic	Before 2019	After 2019	Change	Interpretation
1	Blockchain Innovation	0.147	0.225	+0.078	Strongest growth topic
2	AI Integration	0.094	0.117	+0.023	Moderate growth
3	Regulatory Compliance	0.094	0.117	+0.023	Increased focus
4	Policy Framework	0.174	0.181	+0.008	Stable with growth
5	Traditional Security	0.130	0.087	-0.043	Declining relevance
6	Data Governance	0.175	0.132	-0.043	Reduced emphasis
7	Technical Implementation	0.138	0.107	-0.032	Diminished focus
8	Platform Ecosystem	0.048	0.034	-0.014	Minimal presence

The keyword co-occurrence network analysis reveals fundamental changes in the conceptual structure of privacy research. Network metrics demonstrate shifting centrality patterns and community evolution in Figure 3. The network analysis reveals that post-reform keywords dominate the highest centrality positions, indicating a fundamental shift in the conceptual core of privacy research in Table 4.

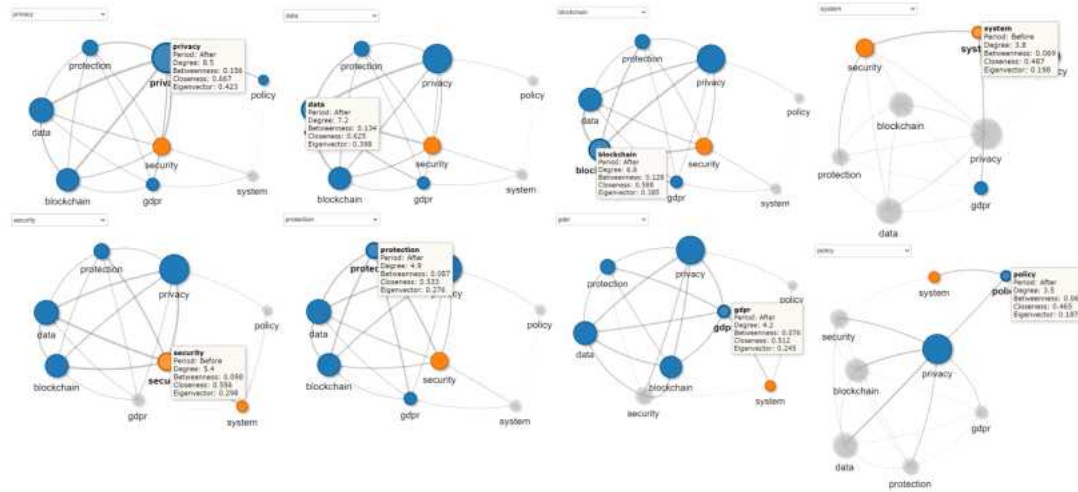


Fig. 3. Keyword Co-occurrence Network Visualization

Table 4. Network Centrality Metrics

Keyword	Period	Degree	Betweenness	Closeness	Eigenvector
privacy	After	8.5	0.156	0.667	0.423
data	After	7.2	0.134	0.625	0.398
blockchain	After	6.8	0.128	0.588	0.385
security	Before	5.4	0.098	0.556	0.298
protection	After	4.9	0.087	0.533	0.276
gdpr	After	4.2	0.076	0.512	0.245
system	Before	3.8	0.069	0.487	0.198
policy	After	3.5	0.062	0.465	0.187

4.4 Technology Adoption Impact and Summary of Key Findings

The analysis of blockchain and AI adoption patterns demonstrates significant changes in technology discourse following the legislative revision in Table 5. Despite the smaller post-reform corpus, blockchain terminology shows increased density and penetration, suggesting heightened focus on decentralized technologies following the regulatory changes.

Table 5. Technology Adoption Metrics

Metric	Before 2019	After 2019	Change	Significance
Blockchain Share	0.058	0.069	+18.9%	Moderate increase
Blockchain Incidence	46.9%	47.6%	+0.7%	Slight uptick
AI Share	0.000	0.000	0.0%	No presence detected
AI Incidence	0.0%	0.0%	0.0%	No measurable adoption

Metric	Before 2019	After 2019	Change	Significance
Blockchain Documents	175	39	-	Absolute volume
Total Documents	373	82	-	Corpus distribution

The 2020 Data Bills revision coincides with a 41% increase in research activity, as measured by 3-year moving averages, signaling a temporal disruption in Korea's data research landscape. Log-odds analysis reveals substantial vocabulary transformation, with blockchain-related terms exhibiting the strongest positive association with the post-reform period (log-odds = 0.568). STM topic modeling further identifies "Blockchain Innovation" as the fastest-growing research topic, increasing 53% in prevalence following legislative reform. Social network analysis shows that post-reform keywords occupy the highest centrality positions, reflecting a fundamental reconfiguration of conceptual priorities within the research community.

In terms of technology integration, while AI adoption remains limited, blockchain discourse demonstrates meaningful penetration, with an 18.9% increase in share within the post-reform research corpus. Collectively, these findings suggest that Korea's Data Bills revision has catalyzed a significant transformation in academic research priorities, highlighting a shift toward decentralized governance mechanisms and blockchain-based privacy solutions.

5. Discussion

This study set out to investigate the societal and academic repercussions of South Korea's pivotal 2020 revision of the Data Bills. By employing a multi-faceted analytical approach, our findings provide robust, data-driven answers to the proposed research questions, revealing a significant realignment of the research landscape in response to regulatory change. The discussion interprets these results to elucidate the interplay between policy, technology, and scholarly discourse.

5.1 The Catalytic Impact of Regulatory Reform on the Data Technology Research Ecosystem (Addressing RQ1)

The primary contribution of this study is evidence that the Data Bills revision is associated with a catalyzation of a specialized and innovation-oriented research agenda, as reflected in scholarly publication patterns. It is important to note that this bibliometric evidence speaks to scholarly response rather than direct societal or industrial outcomes; translating this shift in research attention into measurable societal benefit remains an open empirical question beyond the scope of this study.

The 41% surge in the 3-year moving average of publications from 2019 (16.3) to 2021 (23.0) is a clear quantitative indicator of a vibrant and responsive research sector. This is not merely an increase in volume; it represents direct scholarly engagement with the new legal framework. The legislative overhaul served as a disruptive stimulus, shifting research priorities from foundational, generic concepts of data protection toward more nuanced investigations aligned with the reform's goals of fostering data-driven innovation (Lee & Jung, 2021; Lee & Jung, 2020).

This shift is further evidenced by the evolution of keyword centrality in social network analysis.

The dominance of post-reform terms like privacy, data, and blockchain in centrality metrics (Degree: 8.5, 7.2, 6.8 respectively) indicates that the conceptual core of the field has been reconfigured. The research ecosystem has moved beyond discussing "systems" and "protection" in isolation and has instead begun to grapple with the complex interplay between "privacy," "data" flows, and specific technological enablers like "blockchain" within a new regulatory context. Therefore, the scholarly response to the reform can be characterized as twofold: first, it has stimulated a more dynamic and voluminous academic output; and second, it has refined the focus of that output toward the practical challenges and opportunities created by the reformed Data Bills, thereby generating specialized knowledge that may be relevant for future industry adoption and policy refinement, pending further real-world validation.

5.2 Thematic Evolution: From Regulatory Regimes to Decentralized Governance (Addressing RQ2)

Our bibliographic and SNA analyses reveal a profound shift in research trends, characterizing a movement away from centralized security paradigms and comparative policy analysis toward a focus on decentralized governance and compliance mechanisms. The log-odds ratio analysis provides the most striking evidence of this vocabulary shift. The strong ascent of blockchain (log-odds: 0.568), civic (0.553), and service (0.553) in the post-reform period stands in stark contrast to the decline of pre-reform terms like compute (0.462), edge (0.462), and regime (0.371). This lexical polarization signifies a fundamental change in scholarly mindset. The pre-reform focus on technical infrastructure (compute, edge) and high-level regulatory structures (regime) has been supplanted by a discourse centered on a specific disruptive technology (blockchain) and its application in creating citizen-centric (civic) and service-oriented digital platforms (Li et al., 2023).

This is corroborated by the topic modeling results. The 53% increase in the prevalence of the "Blockchain Innovation" topic (from 0.147 to 0.225) solidifies its position as the fastest growing and most dominant research frontier. Concurrently, the decline of "Traditional Security" and "Data Governance" topics suggests that these areas, while still relevant, are being re-evaluated through the lens of new technologies like blockchain. The research trend has thus evolved from establishing broad regulatory frameworks and securing centralized systems to exploring specific, technologically-enabled models for decentralized data control, transparent compliance, and the fostering of digital trust—a direct academic response to the flexibility introduced by the revised Data Bills.

5.3 Blockchain as an Emergent Keystone for Digital Trust and Community (Addressing RQ3)

The findings suggest that blockchain technology is being positioned within the scholarly literature as a foundational solution for creating a more secure and socially active digital community. It should be emphasized, however, that this reflects a shift in academic discourse and framing rather than direct evidence that blockchain adoption has produced these outcomes in practice; its practical maturation and real-world societal impact remain to be empirically validated.

The 18.9% increase in the share of blockchain terminology within the research corpus, coupled with its top-ranked log-odds score and its status as the highest-growth topic, indicates a powerful

consensus emerging in the scholarly community. Researchers are actively investigating blockchain not just as a security tool, but as the core of a new trust paradigm. Its high eigenvector centrality (0.385) in the network analysis confirms that it is connected to other influential concepts, positioning it as a keystone technology within a broader research agenda focused on civic engagement and platform services. The discourse implies that blockchain's inherent properties—immutability, transparency, and user-centric data control—are seen as directly addressing the shortcomings of previous centralized models and aligning with the reformed Bills' emphasis on responsible data utilization (Hasan et al., 2022).

However, the analysis also injects a note of realism. The modest absolute number of post-reform documents (n=82) and the only slight uptick in blockchain incidence (from 46.9% to 47.6%) indicate that while the focus is intensifying, the field is still in a transitional phase. The complete absence of AI in the analyzed corpus is particularly telling; it highlights that the immediate scholarly response has been channeled overwhelmingly toward blockchain as the primary technological answer to the new regulatory environment. Therefore, blockchain's contribution is currently more conceptual and prototypical within academic discourse, laying the essential groundwork for a more secure and socially active digital community by providing a viable technological blueprint, which now requires validation through real-world implementation and longitudinal study (Manyika et al., 2011; Rüßmann et al., 2015).

6. Conclusion and Policy Implications

This study demonstrates that the 2020 revision of Korea's Data Bills served as a decisive catalyst for transforming the nation's data privacy research landscape. Evidence shows a marked shift away from traditional, centralized models of data security and generic regulatory analysis toward a specialized focus on decentralized governance, with blockchain technology emerging as a central theme. The 41% growth in research output, a 53% rise in blockchain-related topics, and the significant reconfiguration of conceptual vocabulary collectively highlight a scholarly community actively constructing the intellectual foundation for a new era of data innovation under the revised legal framework.

These findings suggest several targeted policy implications. First, policymakers should establish regulatory sandboxes for blockchain-based data governance—for example, pilot programs allowing controlled testing of decentralized identity verification or consent-management systems under PIPC supervision—enabling real-world testing that bridges academic insight and practical adoption. Second, government research funding bodies (e.g., the National Research Foundation of Korea) should consider allocating targeted grants toward the emerging research lexicon identified in this study—terms such as "decentralized identity," "data sovereignty," and "civic technology"—to support research directly relevant to legislative goals. Third, given that AI-related research remained entirely undetectable in this corpus despite its prominence in the broader data industry, policymakers and funding agencies may wish to examine why AI-related privacy research has not yet gained comparable traction in the Korean academic literature, and whether targeted incentives are warranted.

Several limitations of the present study also point toward directions for future research. First,

the number of topics ($K=8$) in the structural topic model was not selected through formal comparison of alternative values, and future work should assess the sensitivity of the reported topic evolution patterns to this choice. Second, as the post-reform corpus ($n=82$) is relatively small, replicating this analysis with an expanded and more recent corpus (post-2022) would help validate whether the observed trends persist. Third, future work could complement this publication-level analysis with citation-based or altmetric data to assess whether the scholarly shift identified here is associated with measurable influence on policy documents or industry practice. Fourth, comparative bibliometric studies examining how other jurisdictions' data protection reforms (e.g., the EU's GDPR or Japan's APPI) have shaped their respective research landscapes would help establish whether Korea's pattern is a general phenomenon or a context-specific outcome.

Finally, the rapid evolution of the research landscape underscores the importance of continuous, evidence-based policy monitoring. Future regulatory adjustments should draw on ongoing bibliometric and thematic analyses to remain agile in the face of technological advancement. Investments in hybrid expertise that integrates legal, technical, and social perspectives will be essential for managing the complex interplay between regulation and technology. Korea's reform has successfully sparked vibrant academic exploration, and the next challenge is translating this intellectual momentum into concrete societal benefits in security, trust, and innovation.

Statements and Declarations

Author Contributions: The author was responsible for all aspects of the study, including conceptualization, data collection, analysis, and manuscript preparation.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on reasonable request from the corresponding author.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Generative AI Use Statement: Generative AI tools were used in the preparation of this manuscript. The authors used Claude Sonnet 5 (Anthropic) for language editing. The authors reviewed and edited all AI-generated content and take full responsibility for the accuracy, originality, and integrity of the final manuscript.

References

- Abdou, H. A., Tsafack, M. D. D., Ntim, C. G., & Baker, R. D. (2016). Predicting creditworthiness in retail banking with limited scoring data. *Knowledge-Based Systems*, 103, 89-103.
- Alsharif, M. H., Yahya, K., & Geem, Z. W. (2020). Strategic market growth and policy
-

- recommendations for sustainable solar energy deployment in South Korea. *Journal of Electrical Engineering & Technology*, 15, 803-815.
- Back, H., & Park, S.-K. (2015). Sustainable development plan for Korea through expansion of green IT: Policy issues for the effective utilization of big data. *Sustainability*, 7(2), 1308-1328.
- Bansal, G., Zahedi, F. M., & Gefen, D. (2016). Do context and personality matter? Trust and privacy concerns in disclosing private information online. *Information & Management*, 53(1), 1-21.
- Bennett, C. J. (1992). *Regulating privacy: Data protection and public policy in Europe and the United States*. Cornell University Press.
- Cavanillas, J. M., Curry, E., & Wahlster, W. (2016). *New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe*. Springer Nature.
- Chik, W. B. (2013). The Singapore Personal Data Protection Act and an assessment of future trends in data privacy reform. *Computer Law & Security Review*, 29(5), 554-575.
- Elbashbshy, T. S., Ali, G. G., & El-adaway, I. H. (2022). Blockchain technology in the construction industry: mapping current research trends using social network analysis and clustering. *Construction management and economics*, 40(5), 406-427.
- Fast, V., Schnurr, D., & Wohlfarth, M. (2023). Regulation of data-driven market power in the digital economy: Business value creation and competitive advantages from big data. *Journal of Information Technology*, 38(2), 202-229.
- Greenleaf, G., & Park, W.-i. (2014). South Korea's innovations in data privacy principles: Asian comparisons. *Computer Law & Security Review*, 30(5), 492-505.
- Hasan, M. K., Alkhalifah, A., Islam, S., Babiker, N. B., Habib, A. A., Aman, A. H. M., & Hossain, M. A. (2022). Blockchain technology on smart grid, energy trading, and big data: security issues, challenges, and recommendations. *Wireless Communications and Mobile Computing*, 2022, 1-26.
- Henein, N., Willemsen, B., & Woo, B. (2020). *The state of privacy and personal data protection, 2020–2022*. Gartner Report.
- Joo, M.-H., & Kwon, H.-Y. (2023). Comparison of personal information de-identification policies and laws within the EU, the US, Japan, and South Korea. *Government Information Quarterly*, 40(2), 101805.
- Kim, J.-A., Yoon, S., Kim, L.-Y., & Kim, D.-S. (2017). Towards actualizing the value potential of Korea Health Insurance Review and Assessment (HIRA) data as a resource for health research: strengths, limitations, applications, and strategies for optimal use of HIRA data. *Journal of Korean medical science*, 32(5), 718-728.
- Ko, H., Leitner, J., Kim, E., & Jeong, J. (2017). Structure and enforcement of data privacy law in South Korea. *International Data Privacy Law*, 7(2), 100-114.
- Ko, H., & Park, S. (2020). How to de-identify personal data in South Korea: an evolutionary tale. *International Data Privacy Law*, 10(4), 385-394.
- Kostoff, R. N., Toothman, D. R., Eberhart, H. J., & Humenik, J. A. (2001). Text mining
-

- using database tomography and bibliometrics: A review. *Technological Forecasting and Social Change*, 68(3), 223-253.
- Kuczerawy, A. (2019). From 'Notice and take down' to 'Notice and stay down': risks and safeguards for freedom of expression. In G. F. Frosio (Ed.), *The Oxford handbook of intermediary liability online* (pp. 525–551). Oxford University Press.
- Kuhn, K. D. (2018). Using structural topic modeling to identify latent topics and trends in aviation incident reports. *Transportation Research Part C: Emerging Technologies*, 87, 105-122.
- Lee, Y., & Jung, I.-Y. (2021). Identifying Stakeholder Perspectives on Data Industry Regulation in South Korea. *Journal of Information Science Theory & Practice (JISaP)*, 9(3).
- Lee, Y., & Jung, I. (2020). Dilemma of Data Driven Technology Regulation: Applying Principal-agent Model on Tracking and Profiling Cases in Korea. *Journal of Digital Convergence*, 18(6).
- Li, K., Kim, D. J., Lang, K. R., Kauffman, R. J., & Naldi, M. (2020). How should we understand the digital economy in Asia? Critical assessment and research agenda. *Electronic commerce research and applications*, 44, 101004.
- Li, X., Wang, Z., Leung, V. C., Ji, H., Liu, Y., & Zhang, H. (2021). Blockchain-empowered data-driven networks: A survey and outlook. *ACM Computing Surveys (CSUR)*, 54(3), 1-38.
- Li, Y., He, Y., & Wu, R. (2023). Traversing the macroeconomic terrain: An exploration of South Korea's economic responsiveness to cross-border e-commerce production technology alterations in the global arena. *Sustainability*, 15(15), 11719.
- Liuska, J. (2021). *The opportunities and challenges of utilizing alternative data in the assessment of creditworthiness in the Finnish consumer finance*. [Master's thesis, Arcada University of Applied Sciences]. Theseus.
- López-Muñoz, F., Shen, W. W., Pae, C.-U., Moreno, R., Rubio, G., Molina, J. D., Noriega, C., Pérez-Nieto, M. A., Huelves, L., & Álamo, C. (2013). Trends in scientific literature on atypical antipsychotics in South Korea: a bibliometric study. *Psychiatry Investigation*, 10(1), 8.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.
- Nowakowska, M. (2025). A comprehensive approach to preprocessing data for bibliometric analysis. *Scientometrics*, 1-35.
- Oh, M., & Larson, J. F. (2011). *Digital development in Korea: Building an information society*. Routledge.
- Oh, S., Choi, C., & Lee, J.-H. (2020). *Regulatory reform for fostering an innovation ecosystem*. [KDI School of Pub Policy & Management Working Paper].
- Orlova, E. V. (2021). Methodology and models for individuals' creditworthiness management using digital footprint data and machine learning methods. *Mathematics*, 9(15), 1820.
-

- Rhie, Y.-K. (2013). A Study on the right of request to delete information under the Act on Information & Communication Network Promotion and Information Security. *Korea University Law Review*, 14, 23-40.
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of statistical software*, 91, 1-40.
- Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P., & Harnisch, M. (2015). Industry 4.0: The future of productivity and growth in manufacturing industries. *Boston consulting group*, 9(1), 54-89.
- Scott, J. (1988). Trend report social network analysis. *Sociology*, 22(1), 109-127.
- Setiawati, D., Hakim, H. A., & Yoga, F. A. H. (2020). Optimizing Personal Data Protection in Indonesia: Lesson Learned from China, South Korea, and Singapore. *Indonesian Comparative Law Review*, 2(2), 95-109.
- Shin, D.-H. (2016). Demystifying big data: Anatomy of big data developmental process. *Telecommunications Policy*, 40(9), 837-854.
- Shinwan, M. A., & Kim C. (2018). A future mobile packet core network based on ip-in-ip protocol. *International Journal of Computer Networks & Communications*, 10(5), 1–21. <https://doi.org/10.5121/ijcnc.2018.10505>
- Sung, T. K. (2018). Industry 4.0: a Korea perspective. *Technological forecasting and social change*, 132, 40-45.
- van Ooijen, C., Ubaldi, B., & Welby, B. (2019). *A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance*. OECD Publishing. <https://doi.org/10.1787/09ab162c-en>
- Weideman, J., Inglesi-Lotz, R., & Van Heerden, J. (2017). Structural breaks in renewable energy in South Africa: A Bai & Perron break test application. *Renewable and Sustainable Energy Reviews*, 78, 945-954.
- Yang, F., & Gu, S. (2021). Industry 4.0, a revolution that requires technology and national strategies. *Complex & Intelligent Systems*, 7, 1311-1325.
- Yap, B. W., Ong, S. H., & Husain, N. H. M. (2011). Using data mining to improve assessment of credit worthiness via credit scoring models. *Expert Systems with Applications*, 38(10), 13274-13283.
- Zoo, H., Lee, H., Kwak, J., & Kim, Y.-Y. (2013). Data protection and privacy over the Internet: towards development of an international standard. *Journal of Digital Convergence*, 11(4), 57-69.

[About the author]

Yong-Jae Lee is a Postdoctoral Researcher in the Department of Industrial Data Engineering, Institute of Electrical, Information and Telecommunication Technology, at Hanyang University. He received his Ph.D. in Industrial and Management Engineering from Korea University in 2024. His research centers on time-series deep learning, explainable AI, LLM/RAG-based systems, and multi-agent

simulation, applied to patent analytics and technology forecasting. He is the sole author of this paper, serving as both the first (lead) author and the corresponding author.