A Methodological Framework for Constructing Big Data Governance Systems: Postprint

Authors: An Xiaomi, Wang Lili

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] To remedy the deficiency in research on methodologies for constructing big data governance systems and propose a more generalizable methodological framework for building such systems. [Method/Process] Beginning with the concept of methodology, we systematically review the ISO International Organization for Standardization's definitions of methodology and existing research findings, identify the elements and categories of methodology, and construct a methodological framework; based on the given methodological framework, we conduct a mapping analysis of existing big data governance research findings from the perspective of methodological elements including theory, conceptual models, principles and rules, processes and procedures, methods and evaluation criteria; and on the basis of comprehensively integrating the methodological elements for constructing big data governance systems, we revise and supplement the current methodology for building big data governance systems by incorporating the theoretical foundation of the Deming Cycle (PDCA). [Results/Conclusion] We clarify the elements of the current methodology for constructing big data governance systems and their relationships, and propose a conceptual framework for the methodology of building big data governance systems from a comprehensive integration perspective.

Full Text

Preamble

Research on Methodology Framework for Big Data Governance System Building

An Xiaomi[1],[2],[3], Wang Lili[1]
[1]School of Information Resource Management, Renmin University of China, Beijing 100872
[2]Key Laboratory of Data Engineering and Knowledge Engineering (Renmin Uni-

versity of China), Ministry of Education, Beijing 100872
[3]E-government Research Center, Renmin University of China, Beijing 100872

[**Purpose/Significance**] This paper aims to address research gaps regarding insufficient attention to the methodology for big data governance system building and proposes a more generic methodology framework for constructing big data governance systems. [**Method/Process**] Beginning with the concept of methodology, this paper systematically reviews ISO's definition of methodology and existing research findings to identify methodology components and categories, constructing a methodology framework. Based on this framework, it maps and analyzes existing big data governance research from methodological components including theory, conceptual models, principles and rules, processes and procedures, methods, and evaluation criteria. Drawing on the theoretical foundation of the Deming Cycle (PDCA), it then revises and supplements current methodologies for big data governance system building. [**Result/Conclusion**] The paper clarifies the components and relationships of current methodologies for big data governance system building and proposes a conceptual framework for such methodologies from an integrated synthesis perspective.

Classification Number: G251
Keywords: big data governance; big data governance system; methodology; methodology framework; methodology components
DOI: 10.13266/j.issn.0252-3116.2019.24.005

1 Problem Statement

In the era of big data, the "digital space" formed through datafication increasingly approximates human social and physical spaces. Datafication has become a recognized new paradigm for understanding society and social behavior, as exemplified by the objective quantification and potential tracking of human behaviors and social activities through network media technologies [1]. Big data development significantly impacts industrial upgrading, scientific research, and human cognitive paradigms [2-3], with the inherent value of big data making it a focal point of international competition, particularly in advancing the modernization of national governance systems and capabilities.

Current research on big data governance frameworks and systems has produced numerous approaches, primarily in three areas: (1) Component-based frameworks, such as the widely recognized systems comprising governance objectives, subjects, objects, and processes (activities) [4-5], or frameworks based on information governance, industry, functional scenarios, and data types [6]; (2) Multilayer architectures spanning from data sources through data fusion and knowledge discovery to decision-making layers [11]; (3) Macro-meso-micro structure-based frameworks [10]; and (4) Governance frameworks formed through the deep integration of people, objects, data, and technology [9]. Other scholars have conceptualized governance objects and problem-solving through core concepts [7-8], or proposed governance models covering the entire big data lifecycle (creation, maintenance, utilization) [12]. Additional studies have examined governance mechanisms [13], rule systems [14], actor network-based models [15], and enterprise data asset management methodologies [18]. Some have employed case study methods, literature reviews, expert consultation, and interpretive structural modeling for healthcare big data governance frameworks [19], while others have proposed three-step methodology roadmaps [20].

However, existing research exhibits limitations: (1) Big data governance systems from different perspectives often lack unified theoretical foundations and conceptual systems, with insufficient attention to interoperability and mutual recognition across frameworks. Core concepts are understood at three independent levels: macro-level multidimensional conceptual systems, meso-level management mechanisms, and micro-level organizational strategies [10,14]; (2) While some studies mention methodologies, they fail to clearly define the concept or provide cross-perspective integrated frameworks. Complex systems theory and EDM (Evaluate-Direct-Monitor) models are occasionally referenced, but methodological discussions remain fragmented.

Thus, this paper adopts a synthetic integration perspective to: (1) clarify generic methodology concepts and components; (2) identify methodological elements across different research contexts; and (3) propose a more universal, cross-domain methodology framework for big data governance system building.

2 Defining Methodology Concepts

2.1 Methodology Definition and Components

Methodology is often confused with epistemology or philosophy of science. At its highest level, it represents "the theory of fundamental methods for understanding and transforming the world" [23]—a meta-theory comprising basic principles and philosophical assumptions underlying any research [24]. However, big data governance is a "socio-technical construct" requiring general scientific methodology applicable across disciplines.

Based on ISO definitions and cross-disciplinary research, methodology is consistently associated with research paradigms and theoretical frameworks [25-27] while addressing specific epistemological and logical issues. Crucially, methodology is not equivalent to specific methods but rather describes, explains, and justifies them [22,28]. It prescribes what should be done, in what sequence, and how to achieve optimal results, providing principles and rules for action, whereas methods merely provide tools and techniques [21]. Methodology can be instantiated through specific elements and steps [26,28].

Analysis of ISO definitions reveals six essential components: (1) theoretical foundation; (2) conceptual model; (3) principles and rules; (4) processes and procedures; (5) methods; and (6) evaluation criteria. These form the basis for our methodology component analysis framework [Figure 1: see original paper].

presents definitions and components from various ISO standards and studies, showing consistent elements including principles, procedures, processes, instructions, standards, methods, guidelines, and evaluation criteria.

2.2 Methodology Component Analysis Framework

Synthesizing ISO definitions, methodology must contain: (1) theoretical foundation providing guidance; (2) conceptual model revealing analytical frameworks; (3) principles and rules defining appropriate actions; (4) processes and procedures with direction and sequence; (5) specific methods; and (6) evaluation criteria based on objectives. The relationships among these components form the methodology component analysis framework [Figure 1: see original paper].

3 Mapping and Deconstruction of Big Data Governance System Building Methodology Components

3.1 Literature Survey and Coding Analysis

We conducted comprehensive literature searches in CNKI (285 Chinese articles) and WoS, JSTOR, Springer, and Emerald (123 English articles after deduplication) through January 13, 2019. From 36 representative articles (24 Chinese, 12 English) selected based on relevance to methodology, we performed coding analysis using our component framework. The coding scheme included: ID, system composition, methodology perspective, methodology components, and specific descriptions .

3.2 Methodology Components and Deconstruction

Current research primarily focuses on theoretical foundations, conceptual models, processes and procedures, and implementation methods, while rarely addressing principles/rules or evaluation criteria .

Methodology Perspectives: Four main control perspectives emerge: (1) governance element control [4-9,15-17,19-20,36-44]; (2) governance process control [11-12,45-47]; (3) governance level control [10,48-49]; and (4) governance mechanism control [13].

Theoretical Foundations: Research draws from public governance theories (holistic governance, precision governance, participatory governance) [50], collaboration theory [14,16,51-52], actor-network theory [15], digital governance theory [53-54], information lifecycle theory [12,16], digital continuity theory [14,16], and public value theory [14]. Some employ EDM methodology [17], Petri nets [15], and attribute graph-based process models [47].

Principles and Rules: Key principles include human-centered design emphasizing multi-stakeholder values [9,11,20], data sovereignty and security [40,55], data-driven governance respecting data patterns and ethics [20,53], and alignment with specific scenarios [20,55].

Conceptual Models: Models reveal core concepts through: (1) five-element integration (subject, object, tools, objectives, activities) [9]; (2) macro-meso-micro hierarchy [7-8,10,42]; (3) data-technology-resource perspectives [36]; and (4) data-information-platform-collaboration-security frameworks [45].

Processes and Procedures: Three logical paths dominate: (1) concept definition → motivation analysis → element identification → framework construction → implementation; (2) concept differentiation → definition → motivation analysis → model construction based on existing frameworks; and (3) post-construction case validation, technical solutions, or maturity assessment [5,7,9-10,16,19,36-37,39-40,42-43,45-48].

Methods: Qualitative methods predominate, including literature review, policy analysis, case studies, interviews, comparative studies, and conceptual analysis. Triangulation combining multiple methods is emphasized but underutilized [4-6,10-11,13,19-20,36,41,46-47].

Evaluation Criteria: Largely absent from current research.

4 Integrated Framework for Big Data Governance System Building and Component Relationships

4.1 Proposal of the Integrated Framework

Existing methodologies have complementary strengths but lack evaluation mechanisms and practical implementation pathways. Drawing on PDCA (Plan-Do-Check-Act) as a quality management approach with iterative improvement principles, we propose an integrated framework for big data governance system building that addresses its socio-technical complexity [Figure 2: see original paper].

The integrated framework uses PDCA phases: (1) **Plan:** Understand context, define needs/objectives, and establish conceptual models; (2) **Do:** Build the governance system with multi-theoretical support, following established processes; (3) **Check:** Evaluate theoretical support, goal alignment, and principle adherence; (4) **Act:** Improve and optimize based on evaluation, validating effectiveness.

4.2 Component Relationships in the Integrated Framework

4.2.1 Planning Phase: Addressing Adaptability This phase clarifies specific scenarios and problems, defining governance needs and objectives. The conceptual model should be prioritized, with current mainstream approaches using five dimensions: subject, object, activity, tool, and objective. Multi-stakeholder alliances should be formed, covering cross-level, cross-domain data with lifecycle management and diverse tools (social and technical) to achieve governance goals of capability enhancement, compliance, risk control, and value creation.

4.2.2 Implementation Phase: Addressing Complexity As a complex socio-technical system, big data governance requires multi-theoretical support: public governance theories for subjects, digital governance theory for objects, information lifecycle and digital continuity theories for activities, and public value theory for objectives. Methodology perspectives include element control (most common), process control (lifecycle/business integration), level control (macro-meso-micro), and mechanism control (operation, collaboration, motivation, and security). Implementation should follow established processes while allowing for case validation and technical solutions.

4.2.3 System Evaluation and Improvement Phase: Addressing Long-term Effectiveness These underdeveloped phases require strengthening. Evaluation should employ expert assessment and case validation to examine: theoretical soundness, goal alignment, and principle matching. Human-centered design emphasizes equal participation and multi-value compatibility. Data security and rights principles require establishing data rights as a key governance domain. Data-driven governance reflects resource-based, law-respecting, method-using, and ethics-following principles. Scenario alignment ensures consistency with specific governance concepts and business objectives. Improvement involves optimizing the system based on evaluation results and validating effectiveness under multi-theoretical guidance.

Conclusion

This paper answers three questions: (1) Methodology comprises theory, conceptual models, principles/rules, processes, methods, and evaluation criteria; (2) Current big data governance methodologies lack integrated, multi-perspective frameworks and emphasize elements over principles and evaluation; (3) A PDCA-based integrated framework synthesizes existing components and adds evaluation and continuous improvement.

Critically analyzing existing methodologies through PDCA provides theoretical and practical value for optimizing governance systems, achieving objectives, and adapting to dynamic contexts. Future research will conduct multi-case empirical studies across scenarios to develop best practice benchmarks.

References

[1] DIJCK JV. Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology [J]. Surveillance & Society, 2014, 12(2): 197-208. [2] JIN X, WAH BW, CHENG X, et al. Significance and challenges of big data research [J]. Big Data Research, 2015, 2(2): 59-64. [3] MAYER-SCHÖNBERGER V, CUKIER K. Big Data: A Revolution That Will Transform How We Live, Work, and Think [M]. Boston: Houghton Mifflin Harcourt, 2013. [4] KIM HY, CHO JS. Data governance framework for big data implementation with a case of Korea [C]// IEEE International Congress

on Big Data. New York: IEEE, 2017: 384-391. [5] AN XM, BAI XY, HONG XH. Research on components of government big data governance system: A case study of Guizhou [J]. E-Government, 2019(2): 2-11. [6] SOARES S. Big Data Governance [M]. KUANG B, trans. Beijing: Tsinghua University Press, 2014. [7] ZHENG DQ, HUANG LH, ZHANG CH, et al. Concept and reference architecture of big data governance [J]. R&D Management, 2017, 29(4): 65-72. [8] ZHENG DQ, FAN YJ, PAN R, et al. Analysis of concepts and elements of big data governance [J]. Science and Technology Management Research, 2017, 37(15): 200-205. [9] LIANG ZM. Big data governance: Implications for national governance modernization [J]. Journal of Jishou University (Social Sciences), 2015, 36(2): 34-41. [10] AN XM, GUO MJ, WEI W, et al. Big data governance system: Core concepts, motivations, and implementation paths [J]. Information and Documentation Services, 2018(1): 6-11. [11] JU J, LIU L, FENG Y. Citizen-centered big data analysis-driven governance intelligence framework for smart cities [J]. Telecommunications Policy, 2018, 42(10SI): 881-896. [12] COYNE EM, COYNE JG, WALKER KB. Big data information governance by accountants [J]. International Journal of Accounting and Information Management, 2018, 26(1): 153-170. [13] ZHAI Y. Research on China's big data governance model innovation and development path [J]. E-Government, 2018(8): 12-26. [14] AN XM, SONG Y, GUO MJ, et al. Research conception on building government big data governance rule system [J]. Library and Information Service, 2018, 62(9): 14-20. [15] ZHANG SH, GAO HM, YANG L, et al. Research on big data governance based on actor-network theory and Petri nets [C]// International Conference on Computer Supported Cooperative Work in Design. New York: IEEE, 2017: 372-377. [16] MING X, AN XM, SONG G. Research on data governance framework in smart city context [J]. E-Government, 2018(8): 27-37. [17] YANG L, GAO HM, SONG JD, et al. Research and application of data governance framework in big data environment [J]. Computer Applications and Software, 2017(4): 65-69. [18] CHENG YX. Methodology and practice of data asset management in big data era [J]. Computer Applications and Software, 2018, 35(11): 326-329. [19] LI Q, LAN L, ZENG N, et al. A framework for big data governance to advance RHINs: A case study of China [J]. IEEE Access, 2019, 7: 50330-50338. [20] LEE SU, ZHU L, JEFFERY R. A data governance framework for platform ecosystem process management [C]// Business Process Management Forum. Cham: Springer, 2018: 211-227. [21] LIU SL. Research on legal methodology [J]. Chinese Journal of Law, 2001(3): 42-54. [22] KAPLAN A. The Conduct of Inquiry: Methodology for Behavioral Science [M]. San Francisco: Chandler Publishing Company, 1964. [23] Term Online. Methodology [EB/OL]. [2019-01-19]. http://www.termonline.cn/list.htm?k=方法论. [24] HE MN. Normativity, objectivity and complexity: Multidimensional social science methodology [J]. Journal of East China Normal University (Philosophy and Social Sciences), 2018(3): 32-41. [25] MACKENZIE N, KNIPE S. Research dilemmas: Paradigms, methods and methodology [J]. Issues in Educational Research, 2006, 16(2): 193-205. [26] JONKER J, PENNINK B. The Essence of Research Methodology [M]. Berlin: Springer, 2010. [27] AN XM, CHAO LM, XU ST, et al. Knowledge

Management Methods and Techniques [M]. Nanjing: Nanjing University Press, 2012. [28] KATSICAS SK. Risk Management [M]// VACCA JR. Computer and Information Security Handbook. Massachusetts: Morgan Kaufmann, 2009. [29] ISO/IEC. Information technology—Security techniques—Methodology for IT security evaluation: ISO/IEC 18045-2008 [S/OL]. [2019-03-21]. https://standards.iso.org/ittf/PubliclyAvailableStandards/c046412_{{{ISO}}*{{IEC}}*}{{{18045}}_{{2008}} [30] ISO. Enterprise integration—Framework for enterprise modeling: ISO 19439-2006 [S]. Geneva: ISO, 2006: 5. [31] ISO/IEC. Information technology—Security techniques—Systems security engineering—Capability maturity model (SSE-CMM): ISO/IEC 21827-2008 [S]. Geneva: ISO/IEC, 2008: 4. [32] ISO. Intelligent transport systems—Reference model architecture(s) for the ITS sector—Part 5: Requirements for architecture description in security engineering: ISO 14813-5-2010 [S]. Geneva: ISO, 2010: 20. [33] ISO/IEC. Information technology—Generic digital audio-visual systems—Part 8: Management architecture and protocols: ISO/IEC 16500-8-1999 [S]. Geneva: ISO/IEC, 1999: 3. [34] PEFFERS K, TUUNANEN T, ROTHENBERGER MA, et al. A design science research methodology for information systems research [J]. Journal of Management Information Systems, 2007, 24(3): 45-77. [35] MASSARO M, DUMAY J, GARLATTI A. Public sector knowledge management: A structured literature review [J]. Journal of Knowledge Management, 2015, 19(3): 530-558. [36] SONG Y, AN XM, MA GH. Study on big data governance capabilities of US, UK and Australian governments: Content analysis of big data policies [J]. Information and Documentation Services, 2018(1): 12-18. [37] LIU YN, YAN H, YANG JL, et al. Progress and theoretical framework construction of government governance research in big data context [J]. Library and Information, 2017(1): 87-93. [38] ZHANG YJ, AN XM, ZHANG GQ. Construction of maturity evaluation index system for government big data governance [J]. Information and Documentation Services, 2018(1): 28-32. [39] CHENG GM. Research on big data governance model and maturity assessment [J]. Science and Technology Innovation, 2016(9): 6-7. [40] ZHAO G, WANG S, WANG P. Exploration of big data governance technical solutions for data sovereignty [J]. Cyberspace Security, 2017(Z1): 36-42. [41] SOARES S, DEUTSCH T, HANNA S, et al. Big data governance: A framework to assess maturity [EB/OL]. [2019-03-02]. http://ishare.iask.sina.com.cn/f/65423173.html. [42] WU SP, LI P, ZHANG ZF. Design of data governance framework under government big data environment [J]. E-Government, 2019(2): 45-51. [43] KE C, WANG K. Research and application of enterprise big data governance [C]// International Conference on Computer Science and Application Engineering. New York: ACM, 2018: 1-5. [44] ZHENG ZL. Social governance model innovation from big data perspective [J]. E-Government, 2016(9): 55-60. [45] MENG TG, ZHAO J. Big data-driven intelligent social governance: Theoretical construction and governance system [J]. E-Government, 2018(8): 2-11. [46] FAN LJ, HONG XH, HUANG C, et al. Challenges and countermeasures of government big data governance [J]. Big Data Research, 2016(3): 27-38. [47] DUTTA H. Graph-based data governance model for real-time data ingestion [J]. CSI Transactions on ICT, 2015, 3(2/4): 119-125. [48] ZHANG SM, PENG

XF, HUANG HX. Research on data governance framework for open universities [J]. China Educational Technology, 2018(8): 116-126. [49] TANG Y, YI CL. Discussion on constructing government data governance model [J]. Theory Guide, 2018(7): 68-74. [50] ZHANG HB. Big data-driven social governance [J]. Comparative Economic & Social Systems, 2017(3): 64-73. [51] ESPINOSA JA, ARMOUR F. The big data analytics gold rush: A research framework for coordination and governance [C]// Hawaii International Conference on System Sciences. Los Alamitos: IEEE Computer Society, 2016: 1112-1121. [52] XU YH, WU XL. Framework reconstruction of social collaborative governance in big data era: Based on "subject-mechanism-objective" analysis [J]. Theory Guide, 2018(1): 41-47. [53] MA HY, YANG JH. Public governance transformation driven by big data: Basic logic and action framework [J]. Chinese Public Administration, 2018(12): 42-46. [54] SUN ZF, JIANG N, ZHENG CM. Application of big data in foreign government governance and its implications [J]. Journal of Gansu Administration Institute, 2018(4): 12-21, 126. [55] MALIK P. Governing big data: Principles and practices [J]. IBM Journal of Research and Development, 2013, 57(3): 1:1-1:13. [56] LI ZH. Innovation and selection of local government governance tools under big data background [J]. Journal of Hunan University (Social Sciences), 2018, 32(5): 143-149. [57] AN XM, GUO MJ, WEI W. Research on collaborative innovation community capability building in government information system integration and sharing [J]. Information Studies: Theory & Application, 2019, 42(4): 76-82. [58] AN XM, BAI WL, ZHONG WR. Knowledge management solution design supporting collaborative innovation community capability building [J]. Science & Technology Progress and Policy, 2015, 32(6): 136-141. [59] ZHANG HP, GUO N, TANG XK. Promotion logic and future path: Textual analysis of China's government big data policy [J]. Journal of Intelligence, 2018(3): 152-157, 192.

Author Contributions:

An Xiaomi: Topic formulation, research design, paper structure revision, content modification, finalization.

Wang Lili: Research data collection, paper writing.

Announcement on Further Strengthening Punishment for Academic Misconduct

To promote academic ethics and combat academic misconduct, *Library and Information Service* will strengthen penalties, including rejection of submissions from authors (especially first authors and corresponding authors) who engage in duplicate submission, plagiarism, or fabrication for five years; retraction and public announcement of published papers with misconduct; and permanent blacklisting. We are committed to working with the academic community to maintain integrity and promote healthy development of library and information science research.

Library and Information Service Editorial Office
June 2019

Note: Figure translations are in progress. See original paper for figures.