Data-Driven Analysis of Research Trends in Humanities and Social Sciences: An Empirical Study Based on National Social Science Projects (2010-2019) Postprint

Authors: Zhai Shanshan, Ye Dingling, Xu Xin

Date: 2023-04-01T16:02:48+00:00

Abstract

[Purpose/Significance] Against the backdrop of data becoming a key production factor and fundamental strategic resource, this study reviews the development status of data-related research in the humanities and social sciences, reveals the research landscape of data issues, and provides a window for understanding data-driven humanities and social sciences research. [Method/Process] Taking National Social Science Fund projects related to data from 2010-2019 as the research object, and based on the external and content characteristics of these projects, this study analyzes the institutional and disciplinary distribution, keywords, and thematic clustering of data-related research in the humanities and social sciences, compares the commonalities and differences in research hotspots between National Social Science Fund projects and their outcomes, and detects trends in data-empowered humanities and social sciences research. [Results/Conclusion] Research on data issues in the humanities and social sciences focuses on long-term themes such as big data environments, information behavior and information services, and government data and scientific data research; identifies database construction and system design, media platforms and services, and international issues research as secondary hotspot themes; and shows development trends toward integrated multi-environments, open data management, extended knowledge organization, and smart humanities services.

Full Text

Analysis of Research Trends in the Field of Humanities and Social Sciences Driven by Data: An Empirical Study Based on National Social Science Projects from 2010-2019

Zhai Shanshan[1], **Ye Dingling**[2], **Xu Xin**[2,3]

[1]School of Information Management, Central China Normal University, Wuhan 430079
[2]Faculty of Economics and Management, East China Normal University, Shanghai 200062
[3]Social Survey and Data Center, East China Normal University, Shanghai 200241

Abstract: [Purpose/Significance] Against the backdrop of data becoming a key production factor and basic strategic resource, this paper reviews the current state of data-related research in the humanities and social sciences, reveals research trends on data issues, and provides a window into humanities and social sciences research driven by data. [Method/Process] Taking National Social Science Fund projects related to data from 2010-2019 as the research object, and based on the external and content characteristics of these projects, this study analyzes the institutional and disciplinary distribution, keywords, and thematic clustering of data-related research in the humanities and social sciences. It compares the commonalities and differences in research hotspots between National Social Science Fund projects and their outcomes, and explores research trends in data-enabled humanities and social sciences. [Result/Conclusion] Research on data issues in the humanities and social sciences focuses on long-term themes including the big data environment, information behavior and information services, government data, and scientific data research, with secondary hotspots in database construction and system design, media platforms and services, and international issues research. Future trends include integrated multi-environments, open data management, extended knowledge organization, and smart humanities services.

Keywords: data; humanities and social sciences; National Social Science Fund; research characteristics

1. Current State of Data-Driven Research in Humanities and Social Sciences

As a new logical starting point for humanities and social sciences researchers to understand research problems, data has broken through the traditional constraints of literature-based and knowledge-based exploration, sparking diverse scholarly perspectives on data issues in the field. In terms of basic theoretical research, scholars have primarily considered the value and risks that data thinking introduces to humanities and social sciences research. Ni Wan et al. identified fundamental problems in applying data to humanities and social sciences research, noting that while the combination claims to use total data, it actually uses sample data; that spatiotemporal data mixing affects precision; and that data reflects correlations but lacks causality. Therefore, humanities and social sciences research must fully consider its own characteristics and the role of data when introducing data thinking. Chen Hongru et al. pointed out that

integrating data thinking with humanities and social sciences research requires establishing basic limits: first, data changes research methods and forms but cannot datafy researchers' spiritual worlds; second, while fully revealing and understanding data is necessary, we must guard against data worship; third, while applying data rationally, we must attend to ethical issues like privacy and security; and fourth, while data can improve research precision, precision does not equal scientific rigor. Shi Qin et al. noted contradictions in humanities and social sciences research processes, such as urgent data needs versus insufficient supply, data security management versus privacy infringement risks, and data objectivity versus falsification. They argued that humanities and social sciences research needs to promote open data resources, build data management and supervision mechanisms, and strengthen data literacy training and education.

Regarding key technology applications, scholars have explored how data-related technologies and platforms transform research scenarios and organizational structures in humanities and social sciences. Ma Feicheng noted that in new data-driven scenarios, humanities and social sciences research needs to develop tangible research tools for data acquisition, storage, analysis, and presentation to improve and innovate research through new technical tools and methodological thinking. Wang Xiaoguang argued that humanities and social sciences development must be driven by data resource construction, supported by specialized databases and computing platforms, and achieved through establishing data resource evaluation standards, standardizing database construction processes, developing data model technologies, and employing data analysis tools to manage and construct humanities and social sciences data resources. Gu Jun et al. designed and built a humanities and social sciences data sharing alliance platform using blockchain technology to address weak data traceability and difficult usage tracking, enabling data sharing, traceability, and tracking. Meanwhile, data development applications have been implemented in many universities and research institutions, such as Fudan University's Social Science Data Platform, which aims to acquire, organize, and develop social science data to provide researchers with rich and timely data services, and Nanjing University's Institute for Humanities and Social Science Big Data, which seeks to break information barriers and integrate, develop, and provide data resources and technologies to promote research development and disciplinary construction.

In terms of service system construction, scholars have discussed research innovation and smart service model construction in humanities and social sciences under data-enabled paradigms. Li Yang et al., through analyzing the origins, context, and characteristics of humanities and social sciences data research, argued that such research needs to move from shallow data presentation to intelligent and meaningful knowledge identification, achieving a Chinese-characteristic humanities and social sciences data research system. Mi Jianing et al. believe that data-driven paradigms have advanced the epistemology and methodology of humanities and social sciences research, transforming research thinking and models and building a multi-integrated academic research system, while empha-

sizing ethical issues in data application technology and value mining in new humanities and social sciences scenarios. Liu Yunong et al., analyzing the practical dilemmas and demands of humanities and social sciences research under data-driven models, pointed out that building innovative research and smart services requires both data customization, processing assistance, technical guidance, and smart push services for researchers, as well as institutional guarantees like data security, service staff training, and clear data property rights.

Overall, scholars' research on data-driven humanities and social sciences mainly focuses on three directions: first, considering conceptual changes brought by introducing data thinking and background in ubiquitous knowledge environments; second, examining development opportunities brought by data-related technologies and methods for humanities and social sciences research and building knowledge discovery platforms for various social phenomena; and third, focusing on constructing data ecosystems in humanities and social sciences to promote smart development through systematic services. These three directions progress layer by layer, describing and revealing common problems in integrating data into humanities and social sciences research. Building on existing research, this paper uses bibliometric methods to conduct multi-perspective empirical research on data-related projects in National Social Science Fund projects, and analyzes the knowledge structure of project outcome papers to grasp macro-level development trends and micro-level fine-grained characteristics of data-driven humanities and social sciences research.

2. Research Design

2.1 Research Approach

As the most authoritative and influential project funding and management organization in China's humanities and social sciences field, the National Social Science Fund typically focuses on theoretical frontier issues, policy-oriented issues, and major development issues, reflecting the forward-looking, developmental, and interdisciplinary nature of humanities and social sciences research and revealing its dynamic trends. This study first conducts statistical analysis of the annual number of data-related projects, project categories, institutions, and disciplines in National Social Science Fund projects to overview external characteristics and overall development trends of data-enabled humanities and social sciences research. It then performs keyword analysis and comparative analysis of research hotspots on both projects and outcomes, as keywords precisely condense research content and represent research priorities and hotspots to a certain extent. Comparative analysis of research hotspots reveals development characteristics of data research in humanities and social sciences through comparing expected and actual research content. Finally, it comprehensively examines research themes and knowledge characteristics to detect and explore research growth points in data-driven humanities and social sciences.

2.2 Data Sources

This study's sample data comes from the "Project Query" system under "Project Search" on the National Philosophy and Social Science Work Office website, searching with "data" as the keyword to collect and count various projects related to data from 2010-2019, obtaining 952 valid projects with specific fields including project approval number, category, discipline classification, title, approval time, principal investigator, professional position, institution, and institution type. Meanwhile, CNKI was used to obtain bibliographic information of 5,293 research papers funded by these projects.

2.3 Research Methods

This study primarily employs bibliometric analysis, cluster analysis, and content analysis. Manual review was used to standardize institution names (e.g., unifying "Sun Yat-sen University Library" as "Sun Yat-sen University"). Excel and SPSS 25.0 were used for data summary, statistics, and charting for external characteristic analysis. The ROST CM 6.0 text mining platform performed word segmentation on project titles and research papers, with manual identification to screen and merge keywords (e.g., removing non-discriminative words like "research," "model," "analysis" and merging semantically identical words like "China" and "our country"). Ucinet 6.186 generated keyword co-occurrence matrices, Netdraw 2.084 visualized keyword co-occurrence maps, and Gephi 0.9.2 performed content mining for thematic cluster analysis, enabling content characteristic analysis of research hotspots and trends.

3. Analysis of Data Research Support and Characteristics in Humanities and Social Sciences

3.1 Annual Distribution Analysis

Annual changes in the number of funded projects can reflect academic research development levels and evolution processes, as well as the importance of research to some extent. Figure 1 [Figure 1: see original paper] shows the funding status of data-related projects from 2010-2019.

As shown in Figure 1, research on data issues in National Social Science Fund projects began in 2010 and emerged in 2013. This is directly related to the 2009 proposal of the fourth paradigm of data-intensive scientific research and the widespread application of big data technology in 2013, indicating that data research in humanities and social sciences is closely linked to social development, academic environment, and information technology. Trend fitting of annual projects reveals that the number of funded projects follows an exponential distribution, with slow initial growth followed by six consecutive years of over 100 projects, demonstrating sustained national and researcher attention, steadily increasing emphasis and development, and good development space and growth potential.

Table 1 shows the distribution of project categories by year.

Among all project categories, general projects dominate, followed by youth projects, indicating national encouragement and attention to young and middle-aged scholars who have become the backbone of data research in humanities and social sciences, foreshadowing good development prospects. Major projects have strict controls, high requirements, and are difficult to fund. The relatively large number of major data-related projects reflects both high national emphasis on data research and strong alignment between data research and social needs, while also showing that data research from a humanities and social sciences perspective has significant influence, with gradually deepening research depth and maturing research systems.

3.2 Research Institution Analysis

Analyzing the distribution of institutions in National Social Science Fund projects reveals the distribution of data-related research and reflects institutions with high research capabilities. Among the 952 funded projects, 316 institutions are responsible, averaging 3 projects per institution. However, only 96 institutions have 3 or more projects, indicating an uneven phenomenon of concentration and dispersion in data research from a humanities and social sciences perspective. Table 2 shows statistics for institutions with 10 or more projects.

3.3 Disciplinary Distribution Analysis

The disciplinary distribution characteristics of data-related projects in National Social Science Fund projects result from the interaction of relevant theoretical support and methodological support. Analyzing disciplinary evolution characteristics helps grasp disciplinary construction of data research in humanities and social sciences and detect external development directions and dynamic trends. Figure 2 [Figure 2: see original paper] shows the disciplinary evolution map.

Overall, data-related research in humanities and social sciences presents a flourishing diversity. Locally, library and information science, statistics, and management science stand out prominently in both project numbers and growth, representing the main disciplines of data research. Journalism and communication and linguistics have moderate project numbers with continuous growth, representing potential development disciplines.

As disciplines that acquire, organize, and utilize data resources, extract and mine information, and develop knowledge theories, technical methods, and application tools, library and information science is inherently an important hub for data resources, making data research a necessary and long-term trend. Statistics, rooted in mathematics, has data running through all its research, making data research particularly prominent. Management science, with its breadth and interdisciplinarity, sees data thinking and technical methods gradually changing

social productivity and production relations, triggering lasting changes in management elements, thinking, and boundaries, and driving management research toward long-term transformations brought by data elements. Therefore, these three disciplines are both the current main disciplines and important future supports for data research in humanities and social sciences.

As disciplines that analyze processed secondary information resources and apply data technologies, journalism and communication and linguistics need to long-term expand research elements, deepen disciplinary connotations, and innovate disciplinary systems under data-driven models. Journalism and communication studies the dissemination of news activities and relationships between news and society, relying on mining, utilizing, and serving data resources. Linguistics studies language essence and development laws, requiring analysis of large corpora to derive statistical patterns and explain language system mechanisms and structural evolution. Their development status and long-term requirements indicate they will become major trending disciplines for data research in humanities and social sciences.

4. Analysis of Data Research Priorities in Humanities and Social Sciences

To deeply explore specific research priorities and reveal internal content characteristics of data research, this study identifies research hotspots in National Social Science Fund projects and their funded research papers separately, and conducts comparative analysis.

4.1 Keywords and Thematic Content Analysis of Funded Projects

This study's data source consists of project titles, which generally comprise research perspectives, backgrounds, and content, reflecting core project points. Analyzing keywords from funded projects generated the keyword map shown in Figure 3 [Figure 3: see original paper] (Top 20, frequency ≥ 43). Drawing on dimensionality reduction and clustering algorithms from big data models, thematic cluster analysis through modularity and random algorithms identified research hotspots and directions, as shown in Figure 4 [Figure 4: see original paper].

As shown in Figure 3, data research in humanities and social sciences primarily focuses on keywords like big data, database, China, governance, and innovation. "Big data" appears 492 times with absolute dominance. Since the big data era began in 2013, it has irresistibly transformed production and management models across industries, influencing public consciousness and lifestyles, making it an inevitable key research point. "Database" appears 206 times, indicating that organizing, constructing, and preserving resources remains a major concern. "China" appears 123 times, directly related to National Social Science Fund projects' national-level focus on China's social development issues. "Governance" appears 62 times, related to the Third Plenary Session of the 18th

CPC Central Committee's shift from social management to social governance, showing humanities and social sciences research closely follows national development strategies and emphasizes solving prominent social problems. "Innovation" appears 58 times, indicating that new data-driven development methods and thinking bring new paths for humanities and social sciences research.

For more precise analysis of research priorities, this study used Gephi's community detection for clustering. Gephi employs random algorithms and segmentation modules to achieve group decomposition and aggregation, more directly and optimally decomposing clustering results than Ucinet, with results typically identified through node size and color, where larger nodes indicate stronger keyword association influence, as shown in Figure 4. Through manual review, emergency-related projects all have big data backgrounds and can be merged with big data environment research categories. Thus, data-related fund projects over the past decade can be summarized as: taking the big data environment as an important research background, emphasizing service behaviors like information services and data openness, and emphasizing support behaviors like database construction. This reflects humanities and social sciences' emphasis on integrating social development contexts with scientific research, demonstrating social value, while also focusing on data service behaviors of users, governments, and researchers, demonstrating humanistic value.

(1) Big Data Environment-Based Research. As an important background for contemporary development, big data is both a social phenomenon and an absolute hotspot in humanities and social sciences data research. In-depth project theme analysis reveals that big data fully integrates with humanities and social sciences, using big data thinking to comprehensively deepen research content, solve social development problems, and build disciplinary development systems, highlighting research value. On one hand, the big data era has triggered comprehensive transformations in national strategic development, social early warning governance, social control and emergency response, network information security, resource environment optimization, industrial structure upgrading, disciplinary development innovation, and education method reform. On the other hand, as a technical method, big data's characteristics of multi-source massive, hierarchical complexity, and comprehensive authenticity impact traditional research methods, 催生 ing emerging technologies and methods like data mining, machine learning, and aggregation computing, promoting comprehensive, deep, and innovative humanities and social sciences research. Specifically, analyzing big data-related projects in library and information science from 2013-2019 reveals three stages: first, big data background research on information security and library user services; second, big data technology research on multi-level text mining, semantic mining, public opinion monitoring, and enterprise services; and third, big data ecosystem research on think tank and talent database construction and evaluation, science and education evaluation platform construction, and archive service platform construction. These three stages align with the aforementioned three research directions.

(2) Information Behavior and Information Services Research. Service research meeting user information needs is also a key theme in humanities and social sciences data research, reflecting high attention to people's livelihood issues and effective problem-solving capabilities. Using multi-source data and technical methods to transform audience-centered passive services into user-centered active service models, and then meeting multi-level user needs through personalized recommendation and precise supply constitutes the main content of information services research. Regarding service objects, research includes students, youth, elderly, and enterprises, with diverse and hierarchical user groups effectively reflecting or solving different information service needs. Regarding service subjects, research includes government public services, library smart services, and archive digital services, emphasizing public service system research on predicting user information needs, understanding user behavior, improving service capabilities, and enhancing service quality. Regarding service content, research focuses on knowledge-based smart information services, ontology-based personalized recommendation services, and perception-based fine-grained humanized services, optimizing user service systems, following social development priorities, and demonstrating theoretical and practical value.

(3) Database Construction and System Design Research. Database construction systematically organizes scattered and disordered resources into deep and orderly digital resources, facilitating network dissemination, access, and use, and reproducing and expanding resource value. Related research focuses on characteristic databases to achieve resource sharing and broaden service space. Research content shows diversified characteristics, including static databases like multi-ethnic language vocabularies, intangible cultural heritage, and ancient literature resources; dynamic databases like international relations, user behavior, and overseas Chinese education; and audio databases like voice and video databases. This shows that data technology development provides effective capabilities for resource development, collection, cataloging, retrieval, and interaction, while characteristic database construction in various disciplines lays a solid knowledge-based data foundation for deep humanities and social sciences research development.

(4) Government Data-Related Research. Government data contains rich social and economic value, and activating data resources can fully release government data dividends. Research mainly focuses on three aspects: first, government data openness research, including open system establishment, platform mechanism construction, risk prevention, and protection strategies; second, government data governance research, including governance mechanism exploration, social public opinion governance strategies, and platform and system establishment; and third, government data evaluation research, including integrity construction evaluation and network dissemination power evaluation. This shows government data research focuses on openness, integration, governance, application, and evaluation, with diverse perspectives and content. The emphasis on application effects indicates research has moved from mechanism analysis to practical evaluation, reflecting systematic, pragmatic, and practical

characteristics.

(5) Scientific Data-Related Research. As the foundation for scientific hypotheses, analysis, and theory formation, scientific data largely determines research quality, and its reusability, reanalyzability, and credibility provide expandable value. Scientific data research focuses on discovery, acquisition, understanding, and reuse, specifically including open policy research, open model analysis, correlation mining, citation mechanism implementation, and influencing factor exploration, with open access being the main focus. This indicates scientific data research is still in the exploratory development stage with considerable prospects, where in-depth theoretical foundation excavation, theoretical system construction, management mechanism exploration, and innovative practice application will become future priorities.

4.2 Keywords and Thematic Content Analysis of Project Outcomes

As an important carrier form of National Social Science Fund projects, research papers undergo extensive discussion and in-depth argumentation from topic selection to publication, largely representing in-depth research content and specific directions. Analyzing paper keywords reveals research hotspots and priorities.

As shown in Figure 5 [Figure 5: see original paper], research priorities in papers share some similarities with funded projects, such as "big data" and "China," but also develop new points like "network" and "information," further verifying that research papers both represent and develop project content. "China" appears 625 times, "big data" 500 times, "service" 187 times, and "openness" 141 times, showing that big data environment research, user service research, and data openness promotion remain hot topics. "Network" appears 326 times, "information" 253 times, and "knowledge" and "technology" 136 times each, showing interconnections among information as foundation, knowledge as extension, technology as means, and network as platform, presenting systematic, ubiquitous, and applied characteristics.

Based on Gephi clustering results shown in Figure 6 [Figure 6: see original paper] and manual review, research on big data, information services, and data openness remains frontier themes, while knowledge management, new media platforms, and international issues represent new directions. This shows humanities and social sciences data research expands in practice, emphasizing knowledge transformation for smart services, ubiquitous information dissemination in social media, and national development issues from an international perspective, closely following social development hotspots and highlighting contemporary characteristics.

(1) Knowledge Organization and Management Research. Knowledge, as an internal driver of innovation, is an important force for promoting social development and stimulating research vitality. Humanities and social sciences data research shows a "data-information-knowledge" development trajectory, moving from superficial behaviors like data development and acquisition to deep seman-

tic analysis, knowledge production, knowledge mining, and knowledge sharing, indicating a shift from informatization to knowledge discovery and service satisfaction. Knowledge will become a key research focus.

(2) Media Platform and Services Research. Social media, based on internet technology with user-generated content and multi-directional communication, contains rich multi-source information and huge social, economic, and academic value. Related research mainly focuses on two aspects: first, using user information data and big data technology to study user behavior and information dissemination patterns in ubiquitous information environments; and second, conducting user privacy protection and dynamic monitoring management research under information alienation phenomena. This shows data research emphasizes both mining social media data resources and humanistic concerns for humanized development.

(3) International Issues-Related Research. Using data resources and technology to obtain comprehensive factual information represents a new approach to international issues research. Such research focuses on international relations and international economic issues, including the Belt and Road Initiative, international competitiveness analysis, international public health cooperation, international relations conflict prediction, capital market competition, international capital flow, and international trade relations. In a borderless network environment with increasingly close international exchanges and free flow of various elements, how to seize data resources to create new development opportunities and enhance capabilities is a problem that humanities and social sciences data research must address.

4.3 Comparative Analysis of Research Hotspots

National Social Science Fund projects, as representations of social development issues, have comprehensiveness and generality, while project outcomes, as highly condensed actual research, have extensibility and granularity. Comparing and integrating their research hotspots can reflect both expected research frontiers and actual content applications, largely parsing characteristics and value of data research in humanities and social sciences. Big data background, information services, and data openness are important long-term themes, while database construction, new media platforms, and international issues are secondary hotspot themes. Overall, research hotspots show convergence in key themes and divergence in some themes.

4.3.1 Convergence in Key Research Themes Big data background, information services, and data openness as long-term research themes are emphasized in both National Social Science Fund projects and project outcomes, aligning with major national strategic thinking and macro-policies like big data development strategy, information service strategy, and data openness guidelines. Both also emphasize “human”-serving research, such as consumer behavior research, personal information security research, and scientific data user behav-

ior research under big data backgrounds, demonstrating humanistic care and human-oriented services. As social subjects, humans are data resource users, knowledge culture creators, and innovation development promoters, making meeting user needs, facilitating user actions, and protecting user rights enduring priorities in humanities and social sciences research.

4.3.2 Divergence in Some Research Themes National Social Science Fund projects emphasize database construction research, showing attention to the fundamental and supportive role of data resources. Specialized database construction processes data organization to achieve data collection, facilitating control, retrieval, maintenance, and sharing, and fully realizing data value, laying a resource foundation for humanities and social sciences prosperity in digital space. Project outcomes emphasize new media platform and international issues research, showing attention to social phenomena with strong timeliness and relevance. Since the rise of big data and internet+ new business forms, new media platforms have flourished and become indispensable in users' lives, making exploration of new user behavior characteristics in new information environments a persistent hotspot. As international connections tighten and China proposes policies like "accelerating cultivation of new advantages in international cooperation and competition" and the Belt and Road Initiative, exploring international issues and competitive advantages from international perspectives has become a sustained hotspot.

5. Perspectives on Future Data Research Trends in Humanities and Social Sciences

Reviewing domestic literature on humanities and social sciences data research and combining research hotspots, this study argues that future data research in humanities and social sciences will achieve breakthroughs in research background, objects, content, and systems, showing trends of contextual integration, broad application, resource knowledgeization, and service humanization.

Integrating Era Characteristics and Keeping Pace with the Times. Humanities and social sciences research is deeply influenced by national strategies, policy systems, social environment, and international relations, and must grasp development trends to maintain effectiveness and advancement. Against the backdrop of 5G era arrival, deepening big data technology, and close international competition and cooperation, data achieves ubiquitous aggregation and sharing, and collaborative value becomes increasingly prominent. How to fully realize continuous and deep development of data-driven hotspot research, integrate new concepts and perspectives to explore growth points, and maintain research vitality are persistent issues.

Incorporating Data Control and Open Data Management. Open data enables data flow and sharing while changing knowledge exchange ecology and promoting economic and social development. Open data research is an im-

portant component of data-related research but remains in the exploratory development stage with considerable prospects. How to enhance data sharing awareness, establish open standards, guarantee open mechanisms, improve infrastructure, promote practical applications, and establish evaluation systems are urgent problems.

Emphasizing Data Mining and Extended Knowledge Organization. Data thinking and technology have changed information organization content from digitalization to dataization. Information organization tools have evolved from traditional controlled vocabularies to semantic networks, and content has shifted to knowledge organization systems based on semantic hierarchy changes. In the future, semantic organization, analysis, knowledge mining, and identification will be effective breakthroughs for semantic knowledge organization and management.

Data as Resources, Smart Humanities Services. Realizing scientific research value requires serving humanity and society. Focusing on user needs, behaviors, and services is the value of humanities and social sciences research. Under data-driven paradigms, smart data development and intelligent technology upgrades achieve coupling between wisdom and humanities and resonance between data outcomes and humanistic services, representing an important orientation for data research in humanities and social sciences.

This study analyzes data-related National Social Science Fund projects over the past decade to grasp research support, characteristics, priorities, and trends, forming a holistic understanding. In terms of support and characteristics, data research shows disciplinary and geographical biases. For sustainable development, future efforts should emphasize organic integration and collaborative cooperation across disciplines and regions to promote horizontal penetration and vertical inheritance. In terms of priorities and trends, data research forms a "data-information-knowledge-service" mainline against big data and international environment backgrounds, with rich and full research levels. Data brings vitality to humanities and social sciences research, and coordinating data theory thinking with data application research to promote smart transformation represents an important trend.

References

[1] Central People's Government of the People's Republic of China. *Opinions on Constructing a More Perfect Market-oriented Allocation System and Mechanism for Factors* [EB/OL]. [2021-01-03]. http://www.gov.cn/zhengce/2020-04/10/content_{5500740}.htm.

[2] Ni W, Tang X. Value and paradox of big data application in social science research[J]. Southeast Academic Research, 2017(4): 68-78, 247.

[3] Chen H, Zhao N, Wang W. Basic issues of integrating big data into humanities and social sciences[J]. Social Science Abstracts, 2016(2): 16-18.

[4] Shi Q, Li Y. Dilemmas and countermeasures of data-driven humanities and social sciences research: Based on micro-level data considerations[J]. Library and Information, 2019(1): 17-23.

[5] Ma F. Promoting deep integration of information technology such as big data and artificial intelligence with humanities and social sciences research[J]. Evaluation and Management, 2018, 16(2): 1-5.

[6] Wang X. Strengthening construction and management of humanities and social sciences data resources[EB/OL]. [2021-01-03]. http://theory.people.com.cn/n1/2018/0705/c40531-30127206.html.

[7] Gu J, Xu X. Design and implementation of a humanities and social sciences data sharing model: A case study of alliance chain technology[J]. Journal of the China Society for Scientific and Technical Information, 2019, 38(4): 354-367.

[8] Fudan University Social Science Data Platform[EB/OL]. [2021-03-29]. http://fisr.fudan.edu.cn.

[9] Nanjing University Institute for Humanities and Social Science Big Data[EB/OL]. [2021-01-03]. https://skbigdata.nju.edu.cn/data/.

[10] Li Y, Sun J. Value pursuit of humanities and social sciences big data research[J]. Library and Information, 2019(1): 1-7.

[11] Mi J, Zhang C, Li D, et al. The fourth research paradigm: Transformation of social science research driven by big data[J]. Academia, 2018(2): 11-27.

[12] Liu Y, Shi Q. Research on innovation of humanities and social sciences knowledge services under data-driven paradigms[J]. Library and Information, 2019(1): 24-30.

[13] Wu C, Zhu Q. Academic trends of culture-tourism integration research in China: Based on National Social Science Fund and Ministry of Education Humanities and Social Sciences Fund project data from 2009-2019[J]. Tourism Forum, 2019, 12(6): 64-74.

[14] National Office for Philosophy and Social Sciences[EB/OL]. [2021-01-03]. http://www.nopss.gov.cn.

[15] Su X. Return of informatics and intelligence work in the big data era[J]. Journal of the China Society for Scientific and Technical Information, 2017, 36(4): 331-337.

[16] Chen J, Yin X. Emergence, characteristics, and mission of the fourth-generation management from a paradigm shift perspective[J]. Chinese Journal of Management, 2019, 16(1): 1-8.

[17] Luo Z. Disciplinary connotation of journalism and innovation in journalism theoretical systems[J]. Press Circles, 2019(4): 11-26.

[18] Liang J, Liu H. Interdisciplinary research in linguistics: Language universals, human cognition, and big data[J]. Journal of Zhejiang University (Humanities and Social Sciences), 2016, 46(1): 108-118.

[19] Liao J. Analysis of research hotspots and development trends of informatics in China in recent five years: Based on National Social Science Fund projects from 2014-2018[J]. Information Science, 2020, 38(3): 160-166.

[20] Communique of the Third Plenary Session of the 18th Central Committee of the Communist Party of China[EB/OL]. [2021-01-03]. http://www.xinhuanet.com//politics/2013-11/12/c_{118113455}.htm.

[21] Deng J, Ma X, Bi Q. Comparative study of social network analysis tools Ucinet and Gephi[J]. Information Studies: Theory & Application, 2014, 37(8): 133-138.

[22] Wu J. Promoting open data to support open science[J]. Library Journal, 2018, 37(2): 4-10.

Author Contributions:

Zhai Shanshan: Proposed research ideas and framework;
Ye Dingling: Collected and analyzed data, wrote initial draft;
Xu Xin: Provided revision suggestions and revised final version.

Analysis on the Research Situation of Humanities and Social Sciences Field Driven by Data
—Empirical Research Based on 2010-2019 National Social Science Projects

Zhai Shanshan[1], Ye Dingling[2], Xu Xin[2,3]

[1]School of Information Management, Central China Normal University, Wuhan 430079
[2]Faculty of Economics and Management, East China Normal University, Shanghai 200062
[3]Social Survey and Data Center, East China Normal University, Shanghai 200241

Abstract: [Purpose/significance] Under the background that data has become a key production factor and basic strategic resource, this paper reviews the research status of data-related research in the field of humanities and social sciences, reveals the research trend of data problems, and provides a window for the perspective of humanities and social sciences research driven by data. [Method/process] Taking the National Social Science Fund projects related to data from 2010-2019 as the research object, and based on the external characteristics and content characteristics of the National Social Science Fund projects, this paper analyzes the distribution of institutions and disciplines, keywords and

topic clustering of data-related research in the field of humanities and social sciences, compares the commonness and differences of research hotspots between National Social Science Fund projects and project achievements, and explores the research trends of humanities and social sciences empowered by data. [Result/conclusion] The research on data issues in the field of humanities and social sciences focuses on the long-term themes of big data environment, information behavior and information services, government data and scientific data research, with secondary hot topics of database construction and system design, media platform and services, and international issues research, and the research development trends of integrating multiple environments, open data management, extending knowledge organization, and smart humanities services.

Keywords: data; humanities and social sciences; National Social Science Fund; research characteristics

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