

Postprint: Construction of China's Government Data Open Sharing Policy System

Authors: Huang Ruhua, Wen Fangfang, Huang Wen

Date: 2023-08-26T00:00:00+00:00

Abstract

[Purpose/Significance] Government data opening and sharing has been incorporated into the national big data strategy. This paper aims to propose a policy system for government data opening and sharing in China in response to the national big data strategy. [Method/Process] Using evidence-based policy research methods, this paper constructs a policy system for government data opening and sharing in China. [Results/Conclusion] The policy system for government data opening and sharing in China is divided into three layers: infrastructure layer, data management layer, and government governance layer. The infrastructure layer includes policies on data and its metadata, IT technology policies (including social media policies), data center policies, data platform standard guidelines, policies for interoperable systems, data storage repository specifications, etc.; the data management layer comprises policies on data creation and collection, data processing, data archiving and preservation, data opening and sharing, and data acquisition and utilization; the government governance layer consists of data security policies, data rights and interests policies, public participation policies, data capability policies, organizational support policies, and financial investment policies. This system provides a policy framework for government data opening and sharing and offers direction for in-depth research on subsequent policies.

Full Text

Preamble

Policy System Construction for Open Government Data in China

Huang Ruhua¹, Wen Fangfang¹, Huang Wen²

¹School of Information Management, Wuhan University, Wuhan 430072

²School of Political Science and Public Administration, Wuhan University, Wuhan 430072

Abstract

[Purpose/Significance] Government data opening and sharing has been incorporated into China's national big data strategy. This paper aims to propose a policy system for government data opening and sharing in China in response to this national strategy. **[Method/Process]** This study adopts an evidence-based policy research approach to construct the policy system. **[Result/Conclusion]** The proposed policy system comprises three levels: infrastructure, data management, and government governance. The infrastructure level includes policies on data and metadata, ICT technology (including social media), data centers, data platform standards and guidelines, interoperable systems, and data repository specifications. The data management level encompasses policies on data creation and collection, data processing, data archiving and preservation, data opening and sharing, and data access and utilization. The governance level consists of data security policies, data rights policies, public participation policies, data capacity policies, organizational support policies, and financial investment policies. This system provides a policy framework for government data opening and sharing and points out directions for subsequent in-depth policy research.

Government data refers to data and information produced or held by government or government-controlled entities. In the context of big data, government data is regarded as a national asset suitable for sharing with businesses and society and should be opened to the public. Open data is defined as data that is freely available online, free from technical restrictions on reuse, and provided under an open license that permits unrestricted reuse [1]. Open government data refers to structured, machine-readable datasets generated by government that are available for stakeholders to download, reuse, or redistribute without restrictions from dedicated internet portals [2]. The emerging open government data movement aims to encourage the sharing of information with the public using open standards and machine-readable formats, maximizing the value of public sector data through big data analytics for reuse and development. Open data initiatives are more easily implemented with national encouragement. To promote the reuse of government information and social innovation more effectively, an increasing number of countries and international organizations are providing policy support for open government data. These policies confirm the value of open government data and establish the fundamental position of making data easily reusable through open access in a transparent and non-discriminatory manner.

In 2015, the State Council issued the "Outline for Promoting the Development of Big Data" (hereinafter referred to as the "Outline"), which established "forming a regulatory system and policy framework for the rational opening and sharing of public data resources" as a goal for the next 5-10 years, officially placing China's top-level design for government data opening and sharing on the agenda [3]. According to the latest report from the World Wide Web Foundation's "Open Data Barometer" (4th edition) [7-8], China ranks 71st globally

in open government data. Among 12 countries in East Asia and the Pacific region, China ranks 10th in open data readiness, with government policy scoring the lowest among the four evaluation indicators: government policy, government action, civil society, and private sector. This indicates that compared with open government data policies in many countries worldwide, China's policy development is relatively lagging. Although government agencies should not arbitrarily copy each other's policies due to differences in operational context, mission, processes, and systems, there remains significant potential for mutual learning [5]. As China's open government data is still in the exploratory stage, there are many empirical studies on foreign policy formulation and implementation experiences, mostly analyzing foreign policies from perspectives such as data opening principles, organizational management mechanisms, data management, privacy and data security, policy frameworks, and policy implementation and evaluation, attempting to provide references for China's open government data. For example, project teams have conducted case studies on policies in the UK, US, France, Canada, Denmark, New Zealand, Australia, and Brazil [9-16]; Huang Hua [17] studied US federal government data governance policies from the perspectives of data opening, information disclosure (freedom), personal privacy protection, e-government, information security, and information resource management. Research on China's government data opening and sharing policies has mainly focused on analyzing the direction of policy research [18], the policy environment for open government data [19-20], top-level design [21], guarantee mechanism construction [22], data security [23], and personal privacy policies [24-25]. While these studies have positively contributed to formulating the public data resource opening and sharing policy system proposed in the Outline, no research has yet systematically studied the policy system itself. Therefore, our research question is: What kind of policy system needs to be established for government data opening and sharing in China?

To address this research question, we first employed content analysis to conduct a comprehensive survey of the current status of national-level government data opening and sharing policies in China, identifying existing gaps [26]. We then used categorical analysis to construct a framework for policy issues related to government data opening and sharing, defining and refining these policy problems [27]. On this basis, we further sought policy solutions for these identified problems and constructed a policy system for government data opening and sharing in China.

2. Research Method and Process

2.1 Research Method

The evidence-based medicine (EBM) movement emerged in the UK, US, and Canada in the late 1990s. EBM uses systematic research of "clinical evidence" to guide clinical practice and has had a significant impact on non-clinical fields such as health, psychology, library and information science, and public policy. Policymakers in the public policy domain have been urged to move away

from political ideology and adopt more rational approaches based on “scientific facts,” a methodology known as evidence-based policymaking [28], which emphasizes establishing rigorous and objective evidence to inform policy decisions. Evidence-based policymaking has gradually become recognized as a prospective approach to public policy formulation that helps people make policy decisions by using the most effective evidence as the basis for policy development. The UK Economic and Social Research Council (ESRC) specifically funds the Centre for Evidence-Based Policy and Practice to promote research and practice in evidence-based policy, while the US has enacted the “Evidence-Based Policymaking Commission Act of 2016.” Since this study focuses on providing reference policy solutions for government data opening and sharing, we adopted the evidence-based policy research method, collecting organizational evidence to promote policy formulation.

2.2 Research Process

To avoid misunderstandings in practice, the UK Cabinet Office defined evidence in the “White Paper on Modernising Government” [29] as: expert knowledge; published research findings; existing research; stakeholder consultation; previous policy evaluations; internet resources; consultation outcomes; policy option costs; and results of economic and statistical modeling. In their research, J.A. Jacobs et al. categorized evidence types into systematic reviews, peer-reviewed research papers, conference proceedings, technical reports, cost-benefit analyses, and grey literature (such as government reports), noting that searching for raw data and analytical materials may be necessary to find evidence suitable for one’s own context [30].

To comprehensively collect evidence related to policy solutions, this study searched research papers, policy documents, and expert recommendations. First, we searched relevant papers using academic databases. Chinese databases included CNKI, VIP, and Wanfang, using the search strategy SU= (“open government data” OR “government data opening” OR “government open data”) AND “policy”. Foreign databases included Elsevier, Springer, Wiley, EBSCO, Emerald, Taylor & Francis, JSTOR, ProQuest full-text databases, and the ISI Web of Knowledge index database, using the search strategy SU= (“open government data” OR “open data”) AND (“polic” OR “*strateg*” OR “*guideline**”). We also used references from articles to further enrich and expand the literature base. Second, we searched for open government data-related policy documents and expert opinions through the official websites of international organizations such as the United Nations, European Union, OECD, World Bank, Open Knowledge Foundation, World Wide Web Foundation, and W3C, as well as government websites of the top 10 countries in the Open Data Barometer. Literature relevance was determined by scanning titles and abstracts. After filtering the retrieved documents, we obtained 351 papers, 93 policy documents, and 56 expert opinions (search cutoff date: July 25, 2017).

We referred to C.B. Stetler's evidence hierarchy [31] to evaluate evidence based on accuracy, objectivity, reliability, universality, relevance, accessibility, grounding in reality, and practicality. Based on the quality of retrieved evidence, we classified evidence into six levels (see Table 1): Level I is the strongest evidence, consisting of analyses of valid existing research; Level VI is the weakest evidence, typically non-research expert opinions or stakeholder consultations. Since this study aimed to find solutions to identified policy problems, in the specific evidence-based policy analysis process, we extracted evidence for solutions to various sub-problems of government data opening and sharing by reading research papers, policy documents, and expert opinions. Excel was used as an auxiliary tool to classify and count evidence according to problems. On this basis, we conducted further systematic analysis and synthesis of evidence, summarized generally accepted policy solutions (such as using linked data for open data platforms), compared and evaluated different policy recommendations (such as whether to charge for data reuse), and finally integrated the evidence to draw evidence-based conclusions.

3. Government Data Opening and Sharing Policy System Framework and Content

Based on an understanding of the current status of China's government data opening and sharing policies and an analysis of policy problems, and supported by evidence from academic research, policy documents, and expert recommendations on open government data, we designed a policy system framework for government data opening and sharing in China from a national perspective (see Figure 1 [Figure 1: see original paper]). Open government data requires robust and sustainable data infrastructure and a sound data infrastructure coordination strategy [32]; the quality and reusability of government data necessarily require governments at all levels to implement lifecycle-based management of data, all of which require strong government governance as support. Therefore, based on reliable evidence and rational analysis, we divided the entire policy system into three levels from bottom to top: infrastructure level, data management level, and government governance level. This policy system construction not only considers the movement patterns of data throughout its lifecycle but also takes into account various factors in the open government data process. The formation of a coordinated policy system will help coordinate various factors in China's open government data initiative and will also promote the research and formulation of specific policies.

3.1 Infrastructure Level: Data Infrastructure Policy

Data infrastructure policy includes policies on data and metadata, ICT technology (including social media), data centers, data platform standards and guidelines, interoperable systems, and data repository specifications. In data infrastructure, data is the most basic element and the foundation of government data opening and sharing. Without a solid data and metadata foundation, data

opening and sharing cannot be achieved.

3.1.1 Data Format Standards To promote the global open data movement, international organizations such as the Sunlight Foundation and Open Knowledge Foundation have proposed basic requirements for open data, particularly the Five Star Open Data deployment scheme and the “Eight Principles of Open Government Data,” which have been widely recognized and applied internationally. The “Eight Principles of Open Government Data” require [33] that data must be complete, primary, timely, accessible, machine-readable, non-discriminatory, non-proprietary, and license-free. The US, UK, and other countries have established these basic principles and requirements in their open data acts or policies, stipulating that open data must adopt machine-readable, non-proprietary, reusable formats. Since open data consumers include both ordinary users and data development companies, providing sufficient data formats and the most convenient formats for the majority of data users will ensure that citizens, data companies, and other departments can access relevant data. In 2006, T. Berners-Lee released the Five Star Open Data deployment scheme [34]: (1) make data available on the web under an open license in any format; (2) make it available as structured, machine-readable data (e.g., Excel instead of image scans); (3) use non-proprietary formats (e.g., CSV instead of Excel); (4) use W3C open standards (RDF and SPARQL) and URIs to identify things; (5) link data to other data to provide context. Since higher-level data formats mean higher openness of data portals, governments should advocate adopting W3C-recommended open data standards and tools, such as XML, RDF, and JSON formats, to facilitate data exchange and integration.

3.1.2 Data Quality Policy Data quality is relative to data users rather than objective, but it includes objective aspects of data itself and data activity processes. From an objective perspective, data quality mainly depends on data attributes such as accuracy, completeness, consistency, validity, timeliness, and accessibility. ISO/TS 8000-150:2011 “Master Data: Quality Management Framework” [35] specifies the basic principles of master data quality management and requirements for implementation, data exchange, and provenance. We recommend implementing comprehensive data quality management according to ISO international standards and formulating a “Data Quality Assurance Outline” for full-process management of various factors involved in the lifecycle. The policy should include: (1) developing open government data quality plans or guidelines; (2) establishing an open data quality working group for overall coordination; (3) establishing procedures for editing data and ensuring data integrity and consistency; (4) evaluating data quality and conducting data audits; (5) cleaning data; (6) monitoring data; (7) utilizing user feedback; (8) regularly conducting data quality activities; (9) identifying and handling data falsification.

3.1.3 Metadata Policy Metadata is a key enabler for effectively using open data, facilitating user access and understanding. Establishing a unified national metadata standard that complies with international standards can not only simplify coordination of open data across central and local governments and departments but also facilitate integration of open data at the national and even global levels. Different government departments have different metadata, and establishing a complete metadata system requires substantial human, material, and financial resources. Therefore, we recommend: (1) developing metadata that meets user needs, establishing metadata fields according to different user groups for different data types; (2) clarifying metadata formats, encouraging the provision of machine-readable, structured metadata formats, and even open linked metadata; (3) using controlled vocabularies (such as DCAT) to improve metadata quality; (4) publishing metadata structure guidelines; (5) establishing mappings between metadata following international standards and other standard metadata.

3.2 Data Management Level: Lifecycle-Based Data Resource Management Policy for Open Government Data

To ensure full utilization of government data, it is necessary for governments to implement full-lifecycle data resource management. In 2013, the US M-13-13 Memorandum “Open Data Policy—Managing Information as an Asset” [36] required executive departments and agencies to manage information as an asset throughout its lifecycle and established principles to help manage information effectively at each stage of the information lifecycle.

3.2.1 Data Creation and Collection Policy Data creation and collection are the starting points for government data opening and sharing, and data quality must also be controlled from the source. Policies in this area should include: (1) creating and collecting data in compliance with open government data basic principles and data standards; (2) creating common metadata and extensible metadata simultaneously with data creation; (3) creating and collecting data and metadata according to user needs, considering from the perspective of data opening and reuse what data to create and data quality requirements, with users providing demand feedback to data producers; (4) involving data access providers in data collection to act as a bridge between data creators and users; (5) open data providers providing profiles of their created data to offer background information that metadata cannot provide; (6) optimizing data collection methods, using electronic archiving and data harvesting methods; (7) ensuring uniformity of information sources, using semantic web technology to integrate data from different sources; (8) ensuring information management through open licenses; (9) implementing cross-departmental and cross-border data co-construction, conducting cooperation and planning with data that needs to be jointly created and collected as the link, creating data on the same theme with the same standard format and unified definitions; (10) strengthening quality control in the data creation and collection process.

3.2.2 Data Processing Policy Governments need to process and refine data before publication to meet open data standards and make it easily accessible and reusable by the public. Data processing policies mainly include: (1) classifying and managing data according to data confidentiality levels. The US National Archives will continue to implement the Controlled Unclassified Information program to protect or control the dissemination of specified information; (2) establishing national standards for describing and encoding government data, including lists of acceptable open data formats, guidelines for revising open datasets, and using international metadata models to make data properly describable using common and extensible metadata, such as the Australian Government's Government Technology Interoperability Framework; (3) compiling thesauri suitable for government data opening and sharing, such as the Library and Archives Canada releasing the "Government of Canada Core Subject Thesaurus" (GCCT) in SKOS, XML, and plain text versions to support open data strategy implementation; (4) developing open government data cataloging strategies or methods, using controlled vocabularies to unify data classification, standardize data cataloging, and enable users to connect to actual data resources through catalogs; (5) creating and maintaining government internal data inventories, which should indicate when individual datasets can be publicly released; (6) creating and maintaining public data inventories, which government agencies must list in both human-readable and machine-readable formats on open data portals; (7) incorporating interoperability requirements into core agency processes; (8) establishing information systems that support interoperability; (9) adopting semantic web technologies, particularly linked data technology, to achieve data interoperability and integration. Linked data, based on standardized formats and interfaces using RDF series standards for data exchange and queries, enables distributed SPARQL queries of datasets and information discovery through browsing, forming the foundation for publishing, sharing, and connecting data on the internet. Many national open data portals implement open data using linked open data standards.

3.2.3 Data Archiving and Preservation Policy To ensure long-term access and availability of data, it is necessary to archive data with preservation value and preserve it properly. Data archiving and preservation can be approached from infrastructure, data archiving, content preservation, preservation metadata, and version control: (1) providing guidelines for trusted digital repositories, which should follow the OAIS reference model and TRAC (Trustworthy Repositories Audit & Certification) checklist requirements to ensure long-term integrity and availability of objects; (2) data archiving policies that incorporate government data with preservation value into archival management, clarifying the scope, duration, and methods of data archiving; (3) data content storage policies that specify concrete maintenance and management activities after data archiving, including determining data preservation formats, duration, data appraisal, retention, deletion, withdrawal, and other specific matters; (4) providing data preservation procedures; (5) providing preservation metadata.

“Preservation Metadata: Implementation Strategies” (PREMIS) is a specification for preservation metadata management in digital asset preservation and can be used for government data preservation; (6) formulating government archives digitization policies. Although current government information is mainly generated and preserved electronically, government departments still hold large amounts of printed archives. To ensure public access, archives needed by the public should be digitized on a large scale, including selection and evaluation of materials for digitization and copyright issues; (7) conducting version control, explaining differences between original and derived datasets to users, conducting continuous version tracking, and documenting all data changes; (8) establishing mirror strategies for data provider portal websites.

3.2.4 Data Opening and Sharing Policy Much data held by government is a national asset suitable for sharing with society, and only through opening can its potential be released. Therefore, open data largely involves government publicly sharing data it collects and owns. Data opening and sharing policies should include: (1) determining which datasets to publish or not publish. There are three approaches to selecting datasets for publication: upgrading data already in the public domain to open data; following international best practices to publish key and high-value datasets defined in the G8 Open Data Charter, Open Data Barometer, and OKFN Open Data Index; and adopting a demand-driven approach. For datasets not to be published, a list should be provided with specific reasons; (2) methods for determining data publication priorities. The Sunlight Foundation provides several methods or factors for determining data publication priorities: various objectives, actors, and events; direct public participation; and costs of publishing individual datasets. Governments should also create processes for user participation to promote and prioritize data publication [36]; (3) following open data standards and requirements, using maximum technical access formats to ensure high quality of published data; (4) publishing bulk data to meet user needs for bulk downloads; (5) designing open government data for specific target groups [39]; (6) specifying methods and tools for data publication, such as using technical and organizational methods, and standardized tools; (7) publishing metadata simultaneously with data; (8) using unique identifiers to improve the quality and accuracy of data analysis; (9) publishing data creation processes simultaneously with data to provide users with background information; (10) optimizing data publication processes. Before publication, compliance review should be conducted from financial, legal, and organizational perspectives. In the early stages of open data, priority can be given to publishing datasets that require no additional processing [40]. A. Zuiderwijk et al. formulated five principles for optimizing open data publication processes [41]: start thinking about data publication from the beginning; develop guidelines, particularly regarding privacy and policy sensitivity; provide decision support by understanding activities of other actors involved in the publication process; make data publication part of routine procedures or schedules; monitor how published data is reused; (11) continuously publishing and updat-

ing data; (12) publishing data under the same open license to enable free reuse; (13) delivering open data to specific target groups; (14) cross-departmental data sharing. Adopting good data standards is an important factor in data sharing. At the same time, develop new data exchange and sharing standards to achieve data integration and interoperability through semantic mediation; (15) code sharing or providing open source. Code used to create government websites, tools, and other online resources can also be provided as valuable open data itself, using open source solutions as much as possible to achieve sharing [38]; (16) consolidating existing cross-level data portals on a legal basis; (17) formulating national data publication plans to release public data in phases and with priorities; (18) establishing a legal framework for data publication.

3.2.5 Data Access and Utilization Policy Data access is a prerequisite for reuse. Governments should ensure that public datasets are easily discoverable. The US, UK, Brazil, and other countries have established information access policies to ensure citizens' right to access government information. Data access policies should include: (1) establishing a national open data platform and compiling guidelines for using the portal; (2) following W3C basic principles for accessible website design, establishing government website accessibility guidelines or standards, and conducting accessibility reviews and reporting; (3) non-discriminatory access. Establishing policies and standards for comprehensive online information and service access for persons with disabilities, such as the US Rehabilitation Act and Telecommunications Act establishing standards for persons with disabilities to access government information; (4) building information systems that support information accessibility, which must be scalable, flexible, and convenient in system design; (5) improving government document/records management; (6) clarifying access methods to ensure data can be obtained in open or machine-readable formats; (7) eliminating restrictions on information access. Public data should not require application or registration, nor should it require detailed user information [42]. Policies can include disclaimers but should not restrict who uses the data or for what purposes [38]; (8) creating APIs for information access. Government agencies can develop APIs or application programming interfaces to allow third parties to automatically search, retrieve, or directly submit information from databases online [38]; (9) creating permanent, continuous data access. Data updates should be documented for public use; (10) providing online data catalogs and metadata; (11) expanding data access through open licenses and technology; (12) making it easier for individuals to access their own information; (13) promoting data access legislation.

Data reuse refers to any commercial or non-commercial use beyond public tasks. To support open data reuse, the EU revised the Public Sector Information (PSI) reuse directive (Directive 2013/37/EU) in 2013 to encourage widespread availability and reuse of public sector information for private or commercial purposes. Most EU member states have also formulated PSI directives to promote data reuse. Public sector information reuse policies mainly include: (1) clarifying data utilization rights and their specific protected objects; (2) establishing stan-

dard terms of use; (3) minimizing restrictions on dataset reuse in terms of use; (4) clarifying data copyright; (5) establishing terms for reusing datasets based on licenses; (6) providing utilization methods; (7) adopting common pricing principles. Data access should be free. The Open Knowledge Foundation states that data should be provided at no more than reasonable reproduction cost. Australia and the US have strong legislation for free (or low-cost) access to government information. According to the PSI directive passed in June 2013, the default charging policy for reusing public sector documents is marginal cost pricing. Given the public goods attribute of government data, we recommend that governments provide public data for free as much as possible to actively encourage open data reuse.

3.3 Government Governance Policy

3.3.1 Data Security Policy Data analysis, integration, and mining pose threats to personal data, commercial secrets, and national security. Digital security policies include: (1) providing policy guidance for data classification and declassification; (2) providing national security protection guidelines; (3) formulating security regulations for open data platforms and applications; (4) establishing a personal data protection framework to protect citizens' personal data security; (5) establishing specialized privacy protection agencies to provide consultation and support for privacy policies and issues in open government data; (6) appointing privacy protection professionals; (7) granting individuals data rights, implementing privacy notices and consent; (8) conducting privacy impact assessments to ensure open government data programs comply with all applicable privacy requirements [43]; (9) incorporating privacy analysis into every stage of the information lifecycle [36]; (10) formulating continuous privacy monitoring strategies to ensure privacy and security controls function properly; (11) establishing policies for privileged user access to sensitive data to prevent privileged users from arbitrarily disclosing and misusing personal data; (12) establishing technical and organizational measures to separate personal data from other data, and taking legal measures when necessary [39].

3.3.2 Data Rights Policy Open data promotes the fundamental right to access government information. However, the information rights paradigm is shifting to a data rights paradigm, and existing information legislation cannot adapt to current open data. Therefore, we recommend: (1) reviewing and amending the "Regulations on Open Government Information". The regulations should include specifying the minimum number of datasets that each government organization should regularly publish [44], clearly specifying specific response times for public sector data requests, and responding to data requests promptly; (2) clarifying copyright ownership of government data; (3) determining data use conditions through copyright law; (4) when government data is protected by copyright, providing access according to license terms and clearly allowing its reuse and dissemination; (5) using CC licenses as the default license for public data reuse; (6) each government using a unified licensing

system that adopts transparent, easy-to-understand, and easy-to-use Creative Commons (CC) licensing methods; (7) selecting compatible or single licenses. Different organizations may use different licenses, which can cause incompatibility, so compatible licenses must be selected or a specific license must be used; (8) clarifying how to use and reuse government and public sector information; (9) minimizing license restrictions, such as eliminating user registration; (10) using simple, standardized license terms; (11) applying CC vocabularies or RDF vocabularies to describe licenses in machine-readable formats for easy acquisition and understanding by applications.

3.3.3 Public Participation Policy A key principle of open government data is participation. Universal participation means everyone can use, reuse, and redistribute public sector information [45]. Participation is driven by public demand, so data should meet user needs, and a benign open data ecosystem should be formed through demand-driven approaches. Policies in this area should include: (1) developing a public participation handbook that provides best practices for public participation and continuously updates them; (2) encouraging citizen participation in government decision-making; (3) continuously expanding public participation in formulating rules and regulations; (4) organizing national open data innovation competitions, such as designing government platforms for competition activities to expand the impact of open innovation activities; (5) soliciting public opinions before data publication, encouraging the use of social media, web conferences, and public consultations to gather opinions; (6) providing unified mechanisms that allow users to provide demands and suggestions for new open datasets, such as customizing user demand information; (7) monitoring the processing of each request; (8) conducting demand assessments to evaluate the level and nature of needs of different user groups and the degree of data needs among different governments, using specific methods such as analyzing received data requests, website monitoring, or public voting on datasets. Comprehensive assessment of various user needs and capabilities can produce good information, data descriptions, format standards, high-quality data, and data management practices that meet different needs [46]; (9) providing organizational support for user feedback, such as planning appropriate outreach activities and supporting dialogue around data, implementing interaction and feedback mechanisms through formal processes, coordination mechanisms, and dedicated staff, and establishing error handling and service frameworks [47]; (10) establishing formal feedback mechanisms connecting data users with data sources [46].

3.3.4 Data Capacity Policy Enhancing data processing and analysis skills across society, such as data collection, analysis, and understanding capabilities, reducing technical staff shortages, and narrowing the digital divide. Policies should provide support for building data capacity across society, mainly including: (1) establishing open government data training programs to improve data capabilities within the civil service; (2) developing plans for citizens' skills

and knowledge in using open data, encouraging the improvement of computer and data literacy for civil servants and ordinary citizens through crowdsourcing; (3) improving IT personnel skills in implementing open data reuse; (4) implementing national curriculum reforms. For example, the UK's "Seizing the Data Opportunity: A Strategy for UK Data Capability" (2013) proposed [48] that starting from September 2014, the national computer curriculum would require students aged 5-16 to master skills needed to build applications and write computer programs, and mathematics curricula would require students to have abilities in statistics, probability, advanced computing, and modeling.

3.3.5 Organizational Support Policy This policy should include: (1) support from high-level political leaders; (2) encouraging political commitment and support at all levels of government; (3) designating lead departments for open data to manage data infrastructure and serve as liaison points for ministries and data users; (4) establishing viable organizational structures. Only through effective decision-making structures can governance be effectively implemented; (5) positioning organizational responsibilities to provide comprehensive support for policies and technologies; (6) establishing open data operation models and cross-agency cooperation mechanisms to coordinate issues, including inter-organizational collaboration models, intergovernmental cooperation models, and government-citizen cooperation models, providing routine operational resources for operating agencies; (7) reaching agreements on standard terms and cross-level, cross-disciplinary use within the information management legal framework [39]; (8) setting overall goals and indicators for project evaluation; (9) managing and developing partnerships with data demanders.

3.3.6 Financial Investment Policy Governments need corresponding fiscal policies to support ICT infrastructure, ensure staff master data skills, develop applications, and support innovative projects. For example, McKinsey suggests that government leaders can invest in talent, tools, and systems, hiring and training staff at different levels to make better use of data for decision-making [49]. In 2015, the US invested \$34 million in funds and in-kind resources to stimulate climate-related data application innovation and support national climate change preparedness [50].

3.3.7 Staffing Policy Staffing policies can include: (1) appointing a central government chief data scientist; (2) establishing positions corresponding to open data, such as chief data officers and chief technology officers; (3) equipping ICT personnel proficient in handling open data; (4) appointing national and departmental data publication coordinators; (5) clarifying roles and responsibilities of different personnel; (6) implementing employee incentive systems.

3.3.8 Social Cooperation Policy The universal participation advocated by open data determines the broad nature of social cooperation, requiring the formulation of open data cooperation strategies. Mainly including: (1) cooper-

ating with national and international open data communities, such as joining OGP (Open Government Partnership) and learning from experiences, knowledge, and best practices of other national government agencies; (2) formulating cooperation agreements between central and local governments on public data; (3) cooperating with users to develop required concepts, tools, and professional knowledge frameworks and rules; (4) cooperating with external data processors.

3.3.9 Cultural Development Policy Open data will cause cultural changes in government agencies and even society as a whole. Forming new open thinking requires policy support to embed an open culture in the governance system. Policies in this area include [5,45]: (1) regular dialogue with users to optimize data provision, develop standards, create new tools, and establish best practices; (2) enhancing public awareness and understanding of provided data; (3) evaluating the impact of open data, such as on political participation and economic development; (4) treating the publication and allowing reuse of government data as a basic government function; (5) establishing a culture that incorporates open data into daily work processes; (6) cultivating a positive data advocacy culture and public participation culture.

Conclusion

This article adopts an evidence-based policy approach to construct a policy system for government data opening and sharing in China. The policy system is divided into three levels: infrastructure, data management, and government governance. The infrastructure level includes policies on data and metadata, IT technology (including social media), data centers, data platform standards and guidelines, interoperable systems, and data repository specifications. The data management level includes policies on data creation and collection, data processing, data archiving and preservation, data opening and sharing, and data access and utilization. The governance level consists of data security policies, data rights policies, public participation policies, data capacity policies, organizational support policies, and financial investment policies. The article proposes a policy framework for government data and metadata and provides overall policy solutions for data resource management throughout the open government data lifecycle. Based on this, we will continue more in-depth research to further refine the specific content under the policy framework. Since open data is an emerging form of government governance, many related policies on copyright, privacy, data reuse, and other aspects are still being explored internationally and require further research. However, we firmly believe that with the advancement of research on government data opening and sharing, China's practice of data opening and sharing will enter a new stage of development.

References

- [1] Open Knowledge Foundation. Open data handbook version 1.0.0 [EB/OL]. [2017-10-02]. <http://opendatahandbook.org/pdf/OpenDataHandbook.pdf>.

- [2] United Nations. Guidelines on open government data for citizen engagement [EB/OL]. [2017-10-02]. <http://www.gaportal.org/resources/detail/guidelines-on-open-government-data-for-citizen-engagement>.
- [3] State Council. Notice on issuing the outline for promoting the development of big data [EB/OL]. [2017-10-02]. http://www.gov.cn/zhengce/content/2015-09/05/content_{10137}.htm.
- [4] ZUIDERWIJK A, JANSSEN M. A comparison of open data policies and their implementation in two Dutch ministries [C] // BERTOT J, LUNA-REYES LF, MELLOULI S. Proceedings of the 13th annual international conference on digital government research. New York: ACM, 2012: 84-89.
- [5] ZUIDERWIJK A, JANSSEN M. Open data policies, their implementation and impact: a framework for comparison [J]. Government information quarterly, 2014, 31(1): 17-29.
- [6] BERTOT JC, CHOI H. Big data and e-government: issues, policies, and recommendations [C] // MELLOULI S, LUNA-REYES LF, ZHANG J. Proceedings of the 14th annual international conference on digital government research. New York: ACM, 2013: 1-10.
- [7] World Wide Web Foundation. Open Data Barometer 4th edition—global report [EB/OL]. [2017-10-10]. <https://opendatabarometer.org/4thedition/report/>.
- [8] World Wide Web Foundation. Open Data Barometer 4th edition—East Asia and the Pacific regional report [EB/OL]. [2017-10-10]. <https://opendatabarometer.org/4thedition/regional-snapshot/east-asia-pacific/>.
- [9] Huang Ruhua, Liu Long. Policy and legal guarantees for UK government data opening and their implications for China [J]. Library and information service, 2017(1): 1-9.
- [10] Cai Jingxuan, Huang Ruhua. Policy and legal guarantees for US government data opening and their implications for China [J]. Library and information service, 2017(1): 10-17.
- [11] Huang Ruhua, Lin Yan. Policy and legal guarantees for French government data opening and sharing and their implications for China [J]. Library, 2017(3): 1-6.
- [12] Hu Yifang, Lin Yan. Policy and legal guarantees for Canadian government data opening and their implications for China [J]. E-government, 2017(5): 2-10.
- [13] Huang Yuting, Huang Ruhua. Policy and legal guarantees for Danish government data opening and their implications for China [J]. Library and information service, 2017(1): 27-36.
- [14] Cheng Yingui, Lai Tong. Policy and legal guarantees for New Zealand government data opening and their implications for China [J]. Library and information service, 2016, 60(19): 15-23.
- [15] Chen Meng. Policy and legal guarantees for Australian government data opening and their implications for China [J]. Library and information service, 2017(1): 18-26.
- [16] He Naidong, Huang Ruhua. Characteristics of Brazilian government data opening and their implications for China [J]. Library and information service, 2017(1): 37-44.
- [17] Huang Huang. Data governance in the US federal government: policy and

- structure [J]. Chinese public administration, 2017(8): 47-56.
- [18] Ma Haiqun, Pan Pan. Analysis of current research status of open data policies at home and abroad and research trend judgment in China [J]. Journal of library science in China, 2015, 41(5): 76-86.
- [19] Ma Haiqun, Wang Hongshuai. SLEPT analysis and strategic deployment of China's government open data strategy [J]. Information science, 2016, 34(3): 3-8.
- [20] Zhao Rundi. Research on China's open government data policy environment from multiple perspectives [J]. E-government, 2016(6): 97-104.
- [21] Chinese Public Administration Society Research Group. Research on top-level design of China's government data opening [J]. Chinese public administration, 2016(11): 6-12.
- [22] Yue Lixin, Liu Wenyun. Research on construction of guarantee mechanisms for China's government data opening [J]. Library and information service, 2016, 60(19): 40-48, 39.
- [23] Wang Bengang, Ma Haiqun. Policy analysis of open data security issues [J]. Information studies: theory & application, 2016, 39(9): 25-29.
- [24] Zhang Xiaojuan, Wang Wenqiang, Tang Changle. Research on policies and regulations for government data opening and personal privacy protection in China and the US [J]. Information studies: theory & application, 2016, 39(1): 38-43.
- [25] Di Liya. Research on personal data protection in deep opening of government data [J]. Library, 2016(6): 54-58, 64.
- [26] Huang Ruhua, Wen Fangfang. Policy framework and content of China's government data opening and sharing: content analysis of national-level policy texts [J]. Library and information service, 2017, 61(20): 12-25.
- [27] Huang Ruhua, Wen Fangfang. Construction of policy issues for China's government data opening and sharing [J]. Library and information service, 2017, 61(20): 26-36.
- [28] BEHAGUEL D, TAWIAH C, ROSATO M, et al. Evidence-based policy-making: the implications of globally-applicable research for context-specific problem-solving in developing countries [J]. Social science & medicine, 2009, 69(10): 1539-1546.
- [29] Cabinet Office. Modernising government white paper [EB/OL]. [2017-07-08]. http://hansard.millbanksystems.com/westminster_{hall}/1999/dec/09/modernising-government-white-paper.
- [30] JACOBS JA, JONES E, GABELLA BA, et al. Tools for implementing an evidence-based approach in public health practice [J]. Preventing chronic disease, 2012, 9: E116.
- [31] STETLER CB, BRUNELL M, GIULIANO KK, et al. Evidence-based practice and the role of nursing leadership [J]. Journal of nursing administration, 1998, 28(7-8): 45-53.
- [32] DAWES SS, VIDIASOVA L, PARKHIMOVICH O. Planning and designing open government data programs: an ecosystem approach [J]. Government information quarterly, 2016, 33(1): 15-27.
- [33] Open Government Working Group. Eight principles of open government

- data [EB/OL]. [2017-09-10]. <http://opengovdata.org/>.
- [34] BERNERS-LEE T. Linked data [EB/OL]. [2017-09-26]. <https://www.w3.org/DesignIssues/LinkedData.html>.
- [35] International Organization for Standardization. ISO 8000-150: a framework for data quality management [EB/OL]. [2017-09-26]. <http://www.dpadvantage.co.uk/2014/12/23/iso-8000-150-a-framework-for-data-quality-management/>.
- [36] BURWELL SM, VANROEKEL S, PARK T, et al. Open data policy—managing information as an asset [EB/OL]. [2017-09-26]. <https://project-open-data.cio.gov/policy-memo/>.
- [37] LEE D. Building an open data ecosystem: an Irish experience [C] // ESTEVEZ E, JANSSEN M, BARBOZA LS. Proceedings of the 8th international conference on theory and practice of electronic governance. New York: ACM, 2014: 351-360.
- [38] Sunlight Foundation. Open data policy guideline [EB/OL]. [2017-09-26]. <https://sunlightfoundation.com/opendataguidelines/>.
- [39] Federal Ministry of the Interior. Open government data Germany [EB/OL]. [2017-09-26]. https://cdn1.scrivt.com/fokus/8414f0b740831251/b88e6b34c5e755316eaf47d5aca824c4/SQC_{
- [40] Federal Council. Open government data strategy for Switzerland 2014-2018 [EB/OL]. [2017-09-26]. <https://www.e-government.ch/en/umsetzung/e-government-schweiz-2008-2015/open-government-data-schweiz/>.
- [41] ZUIDERWIJK A, JANSSEN M, CHOENNIS, et al. Design principles for improving the process of publishing open data [J]. *Transforming government: people, process and policy*, 2014, 8(2): 185-204.
- [42] Public Sector Transparency Board. Public data principles [EB/OL]. [2017-09-26]. <https://data.gov.uk/library/public-data-principles>.
- [43] Open Government Partnership. Third open government national action plan for the United States of America [EB/OL]. [2017-07-20]. <https://www.data.gov/meta/open-government-national-action-plan/>.
- [44] NUGROHO RP, ZUIDERWIJK A, JANSSEN M, et al. A comparison of national open data policies: lessons learned [J]. *Transforming government: people, process & policy*, 2015, 9(3): 286-308.
- [45] UBALDI B. Open government data: towards empirical analysis of open government data initiatives [EB/OL]. [2017-07-20]. <http://www.oecd-library.org/governance/open-government-data{5k46bj4f03s7}-en>.
- [46] DAWES SS, HELBIG N. Information strategies for open government: challenges and prospects for deriving public value from government transparency [C] // WIMMER MA, CHAPPELET JL, JANSSEN M, et al. Proceedings of the 9th IFIP WG 8.5 international conference on electronic government. New York: ACM, 2010: 50-60.
- [47] SUSH A I, GRONLUND A, JANSSEN M. Organizational measures to stimulate user engagement with open data [J]. *Transforming government: people, process and policy*, 2014, 9(2): 181-206.
- [48] HM Government. Seizing the data opportunity: a strategy for UK data capability [EB/OL]. [2017-09-26]. <https://www.gov.uk/government/publications/uk-data-capability-strategy>.
- [49] CHUI M, FARRELL D, JACKSON K. How government can promote open data [EB/OL]. [2018-02-23]. <https://www.mckinsey.com/industries/public->

sector/our-insights/how-government-can-promote-open-data.

[50] The Open Government Partnership. Third open government national action plan for the United States of America [EB/OL]. [2018-02-23].

https://open.gov/assets/files/final_{{us}}{{open}}{{government}}{{national}}{{action}}{{plan}}}3

Author Contributions:

Huang Ruhua: Guided topic selection;

Wen Fangfang: Determined research methods, collected and analyzed data, and wrote and revised the paper;

Huang Wen: Analyzed Chinese data.

Policy System Construction of Open Government Data in China

Huang Ruhua¹, Wen Fangfang¹, Huang Wen²

¹School of Information Management, Wuhan University, Wuhan 430072

²School of Political Science and Public Administration, Wuhan University, Wuhan 430072

Abstract: [Purpose/significance] Open government data has been incorporated into national strategy in China. This paper aims to propose a policy framework for opening and sharing government data in China. [Method/process] This article adopted the method of evidence-based policy to construct a policy system for opening and sharing government data. [Result/conclusion] The policy system of open government data is divided into three levels: infrastructure level, data management level, and governance level. The infrastructure level includes data and metadata policies, IT technology policies (including social media policies), data center policies, data platform standards and guidelines, interoperable system policies, and data repository specifications. The data management level involves data creation and collection policy, data archiving and preservation policy, policy of opening and sharing data, and policy of access to data and reuse. The governance policy contains data security, data rights, public participation, data capacity, organizational support, financial policy, and so on. The system provides a policy framework for opening and sharing government data and points out a direction for further research on follow-up policies.

Keywords: government data; open government data; open data; data sharing; policy system

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

Source: ChinaXiv — Machine translation. Verify with original.