

---

AI translation · View original & related papers at  
[chinaxiv.org/items/chinaxiv-202308.00500](https://chinaxiv.org/items/chinaxiv-202308.00500)

---

## Evaluation Framework and Post-print Findings for Local Government Open Data

**Authors:** Zheng Lei, Lü Wenzeng

**Date:** 2023-08-27T00:00:00+00:00

### Abstract

[Purpose/Significance] This study constructs an evaluation framework for local government open data in China, evaluates the data dimensions of existing local government open data platforms, and proposes recommendations to promote local government data openness. [Method/Process] Drawing upon the definition, principles, and standards of government open data, referencing international open data evaluation frameworks, and based on current policy requirements and development status of government open data in China while incorporating insights from experts and scholars across various fields, we develop a systematic, scientific, multi-dimensional, and operable government open data evaluation framework, which is then applied to conduct a comprehensive evaluation of data from 46 existing local government open data platforms across China. [Results/Conclusion] The study identifies various issues concerning the quantity, quality, standards, coverage, and sustainability of data opened by local governments in China.

### Full Text

#### Preamble

Vol. 62 No. 22, November 2018. ChinaXiv Partner Journal. Assessment Framework and Findings for Local Government Open Data in China. Zheng Lei, Lü Wenzeng. School of International Relations and Public Affairs, Fudan University, Shanghai 200433.

### Abstract

[Purpose/Significance] This study constructs an assessment framework for evaluating local government open data initiatives in China, examines the data dimensions of existing local government open data platforms, and offers recommendations to promote local government data openness. [Method/Process]

Based on the definitions, principles, and standards of open government data, drawing upon international open data assessment frameworks, and considering China's current policy requirements and development status for government data openness, this research incorporates input from multidisciplinary experts and scholars to build a systematic, multi-dimensional, and operational assessment framework. Using this framework, we conducted a comprehensive evaluation of data available on 46 local government open data platforms in China. **[Result/Conclusion]** The study identifies various issues concerning the quantity, quality, standards, coverage, and sustainability of open data released by Chinese local governments.

**Keywords:** China; local government; open data; assessment; framework

**Classification Number:** G250 **DOI:** 10.13266/j.issn.0252-3116.2018.22.004

Data resources produced, acquired, and preserved by government departments in fulfilling their administrative duties constitute public resources of society. Maximizing the openness of government data to society for development and utilization—while safeguarding state secrets, commercial secrets, and personal privacy—can enhance government transparency, stimulate innovation and entrepreneurship, transform economic development patterns, improve public service levels, and strengthen government governance capabilities [1]. In recent years, as open government data has advanced rapidly worldwide, the Chinese government has attached great importance to this movement. Since 2012, nearly 50 local governments in China have launched open data platforms, achieving certain results and accumulating considerable experience. However, fundamental questions remain: How much data have Chinese local governments actually opened? What are the standards and quality of this data? Which domains does it cover? Is it being opened sustainably? What problems and challenges exist? These questions necessitate in-depth evaluation and research.

This paper first constructs a systematic, multi-dimensional, and operational open data assessment framework. Based on this framework, we conduct a comprehensive evaluation of the data dimensions on existing local government open data platforms in China, propose optimization and enhancement recommendations, and aim to contribute to the advancement and development of local government data openness in China.

## 2 Literature Review

### 2.1 Basic Principles and Standards for Government Data Openness

In December 2007, 30 open data advocates gathered in California and jointly proposed eight fundamental principles for government data openness [2]: complete, primary, timely, accessible, machine-readable, non-discriminatory, non-proprietary, and license-free. According to the World Bank's definition, open data refers to data that can be freely used, reused, and redistributed by anyone for any purpose without restriction, while maintaining its original source

and openness to the greatest extent possible. The Open Definition states that openness means anyone can freely access, use, modify, and share data for any purpose [3]. “Openness” should possess two dimensional characteristics: technical openness, meaning data should be in machine-readable, non-proprietary electronic formats that can be obtained and utilized by anyone using free software, and should be placed on public servers accessible to the public without passwords or firewalls; and legal openness, meaning the data must be in the public domain or under free-use terms with minimal restrictions [4].

In 2010, Tim Berners-Lee, inventor of the World Wide Web and creator and advocate of the Semantic Web and Linked Data, proposed a five-star standard for open data [5]: One star means data is openly licensed and available on the web, allowing users to view, search, store, and modify data, and share it with anyone, with no format requirements (PDF, JPEG, etc. are acceptable). Two stars mean data is provided in a machine-readable, structured format, such as Excel spreadsheets (but not scanned images of tables). Three stars mean data meets two-star standards and is provided in a non-proprietary open format, such as CSV instead of Excel, allowing users to analyze data without requiring proprietary, paid software. Four stars mean data meets all previous requirements and adopts W3C open standards (such as RDF and SPARQL formats), with fixed URL links for each dataset to facilitate discovery and linking to specific dataset locations. Five stars mean data meets all previous requirements and, using W3C standards and linked data principles, enables data interlinking and provides data context.

In 2015, the Open Data Charter defined open data as electronic data possessing necessary technical and legal characteristics to be freely used, reused, and distributed by anyone, anytime, and anywhere. The charter also proposed six principles for government data openness [6]: open by default, timely and comprehensive, accessible and usable, comparable and interoperable, committed to improving governance and citizen engagement, and committed to inclusive development and innovation.

China’s policy requirements for government data openness align with these international standards. In February 2017, the Central Deep Reform Group reviewed and approved the “Several Opinions on Promoting the Opening of Public Information Resources,” which stipulated that open data must possess “completeness, accuracy, originality, machine readability, non-discrimination, and timeliness, facilitating online retrieval, access, and utilization by the public.” In May 2017, the General Office of the State Council issued the “Implementation Plan for the Integration and Sharing of Government Information Systems,” which stated that “original, machine-readable datasets suitable for social reuse” from government departments and public enterprises and institutions should be opened to society. In January 2018, the Cyberspace Administration of China, the National Development and Reform Commission, and the Ministry of Industry and Information Technology jointly issued the “Pilot Work Plan for Opening Public Information Resources,” requiring pilot areas to “improve data complete-

ness, accuracy, validity, and timeliness,” and to “clarify requirements for data completeness, machine readability, and format universality.”

A summary of government data openness standards from relevant domestic and international institutions and documents is presented in Table 1 .

## 2.2 Research on Government Data Openness Assessment

E. Ovieido et al. [7] established an open data platform quality model including six dimensions: usability, reusability, relevance, reliability, granularity, and visualization. G. Viscusi et al. [8] proposed a quality assessment framework for open government data based on three dimensions: completeness, accuracy, and timeliness. R. P. Lourenço [9] proposed seven indicators: data quality, completeness of platform data subjects and time, ease of data access, data usability and understandability, timeliness, data value and usefulness, and granularity. O. Bello et al. [10] used assessment variables including “five-star standard,” implementation technology, data format, open licensing, key datasets, and functionality.

In 2014, the GovLab at New York University reviewed definitions of “open data” from eleven representative research institutions, assessment indicators, government departments, and consulting companies, finding that the most frequently mentioned open data standards included requirements for free access, public provision, non-exclusivity, usable structure, open licensing, and reusability [11]. I. Susha et al. [12] conducted a comparative study of five open government data assessment projects across three dimensions: metadata, meta-methods, and meta-theory.

Chinese scholars have also conducted research on assessment methods for government data openness. Xia Yijian compared and summarized the assessment themes, focuses, objects, and methods of seven international organization open government data assessment projects [13]. Zheng Lei and Guan Wenwen analyzed the assessment frameworks, indicators, and methods of eleven representative domestic and international assessment projects, finding that current open government data assessment indicator systems primarily include five dimensions: foundation, platform, data, usage, and impact, with emphasis on data and foundation levels [14]. Zheng Yueping and Liu Meicen compared several representative foreign open data assessment projects—including the World Bank’s “Open Data Readiness,” the Web Foundation’s “Open Data Barometer,” Open Knowledge Foundation’s “Global Open Data Index,” OECD’s “OURdata Index,” and the UN’s “Open Government Data Survey”—across multiple dimensions such as start time, assessment frequency, assessment objects, assessment tools, data acquisition, and scoring calculation, identifying commonalities and differences and discussing existing problems and deficiencies [15]. Chen Mei analyzed specific practices in open government data value assessment in major countries (United States, United Kingdom, Australia) and international organizations, proposing recommendations to recognize the dilemmas of value

assessment, emphasize and conduct value assessments, and focus on assessment methodologies [16]. Wei Xinling, An Xiaomi, et al. systematically reviewed existing open government data assessment systems, summarized assessment motivations, contents, methods, and characteristics, identified limitations in existing systems regarding city-level, industry-specific, and maturity assessments, and constructed a cyclical iterative testing mechanism for assessment motivations, contents, and methods [17].

Some scholars have conducted actual assessments of China's government data openness status. Zheng Lei and Gao Feng first assessed open data practices in eight Chinese local governments across three levels—foundation, data, and platform—using thirteen dimensions [18]. Zheng Lei and Xiong Jiuyang further studied the technical and legal characteristics of data on thirteen local government data platforms in China, covering dimensions such as quantity, type, format, access method, timeliness, open licensing, metadata, views, and downloads [19]. Xia Yijian systematically analyzed different assessment systems' basic understanding of Chinese government data openness, finding gaps and deficiencies in China's government data openness level, information legal system, organizational management system, and technical architecture [20]. Zhao Jidi and Zhang Hanlun evaluated local government data openness effectiveness using Shanghai as a case study, combining content analysis and interview data across three dimensions: provision and management of open data, citizen participation, and data utilization [21]. Shen Jing et al. established a government data openness development speed assessment system from three dimensions—openness improvement, update frequency improvement, and user utilization improvement—crawling data from five provincial platforms to obtain development speed indices and analyzing trends in combination with China's Open Data Forest Index [22]. Hailun and Deng Song assessed the overall efficiency, pure technical efficiency, and scale efficiency of thirteen city government data platforms in China, finding that nine of the thirteen platforms achieved relatively effective pure technical operation [23].

### 2.3 Review of Current Research

The above review shows that domestic and international scholars have accumulated considerable research on open data assessment methods, and assessments of Chinese government data openness have begun. However, current international assessments of open government data primarily focus on the national level, while existing Chinese assessments of local government data platforms have not been sufficiently comprehensive in sample coverage or sufficiently focused in assessment frameworks. Although they cover data, platform, policy, and management dimensions, assessment indicators specifically targeting the core data dimension are not yet in-depth or systematic. As Chinese local governments continue advancing open data practices with increasingly diverse content and forms, there is an urgent need for more systematic and focused research on a broader range of samples to comprehensively present and deeply analyze the

current status of local government open data in China.

### 3 Evaluation Methodology

Overall, this study follows principles and standards of government data openness, references international open data assessment frameworks and indicator systems, incorporates current Chinese policy requirements and development status, and integrates suggestions from experts and scholars to determine assessment objects, indicator systems, and data collection and analysis methods.

#### 3.1 Assessment Framework

This research focuses on assessing the data dimension of local open data platforms. The observation objects are original electronic datasets and related information publicly available through direct download or API interfaces on local government open data platforms, excluding government data provided to society through internal authorization or agreement-based access rather than public platforms.

This study invited nearly 40 domestic experts and scholars in the data field to jointly construct the assessment framework. These experts and scholars represent diverse disciplinary backgrounds—including public management, information science, computer science, and political science—and come from universities, research institutions, government, and enterprises, reflecting cross-boundary, multidisciplinary, third-party, and neutral professional perspectives and practical needs. First, experts proposed assessment indicators through group discussions based on systematic, scientific, and operational principles. Then, through plenary discussion, they organized and categorized the indicators proposed by each group, merging similar items. Finally, they selected relatively important indicators through on-site voting. This process preliminarily constructed an assessment indicator system for the data dimension of Chinese local government data openness, including primary indicators such as data quality, data standards, data sustainability, data quantity, and data coverage, along with corresponding secondary and tertiary indicators as shown in Table 2 .

Subsequently, experts and scholars anonymously ranked the relative importance of indicators using an online survey tool, ranking the most important indicator as 1, the second most important as 2, and so on. The ranking results are shown in Table 3 . The indicator ranked as most important was “data quality,” reflecting the demand for high-quality datasets across sectors. The indicators ranked 2nd to 5th were data standards, data sustainability, data quantity, and data coverage, respectively.

#### 3.2 Evaluation Objects

Based on public reports and searches using keywords such as “data + openness,” “data + public,” “public + data,” “government + data,” “place name + data,”

“place name + government data,” and “place name + open data,” this study identified government open data platforms launched in China by mid-April 2018. Platforms meeting the following criteria were included in the assessment scope:

- (1) The platform domain contains .gov.cn, confirming it as an official government platform.
- (2) The platform form is either “unified proprietary” or “unified embedded.” “Unified proprietary” means open data is aggregated on a dedicated platform; “unified embedded” means open data is aggregated as a section embedded in a government portal or government service website.
- (3) The platform represents local governments at the prefecture-level city or above.
- (4) The platform actually provides electronic, structured datasets accessible through download or API interfaces. Some platforms named “data openness” essentially only provide unstructured text content or links to other relevant webpages without downloadable or API-accessible structured datasets. These platforms belong more to traditional “information disclosure” portals and were therefore excluded from this assessment, such as the Xinjiang Uygur Autonomous Region Government Data Openness website, the “Open Data” module on the Sichuan Provincial People’s Government website, and the “Data Openness” channel on the Qingyuan Municipal People’s Government website in Guangdong.

Based on these selection criteria, 46 local government open data platforms were included in this study’s assessment. These platforms meet the basic characteristics of government data openness and represent pioneers in China’s government data openness movement. Specific platform names, affiliated local governments, and domain names are shown in Table 4 .

### 3.3 Data Collection and Analysis

This study employed a combination of automated web crawling and manual observation for data collection, with the collection period from April 13 to April 18, 2018. The main assessment analysis was based on data collected from various platforms as of April 18, 2018, while the evaluation of the “dynamic update” indicator used data collected from January 1 to April 18, 2018. The study primarily used descriptive statistical analysis, cross-analysis, and text analysis to analyze the collected indicator data.

## 4 Findings

### 4.1 Data Quantity

**4.1.1 Total Number of Datasets** A dataset is a collection of data, typically appearing in tabular form where each “column” represents a specific variable and each “row” corresponds to a sample unit. Government open data platforms often provide datasets through download or API interfaces. In this study, the

following three types of cases were not considered valid open datasets: (1) No directly downloadable or API-accessible dataset exists under the dataset name; (2) The dataset contains only 0-2 rows of low-capacity data; (3) The dataset name provides only a webpage link that redirects to text content that cannot be obtained through download or interface. Figure 1 [Figure 1: see original paper] shows the top 10 local platforms with the highest total number of effective open datasets (including both direct download and API interface access).

**4.1.2 Data Capacity** Data capacity refers to the total data volume calculated by multiplying the number of fields (columns) by the number of records (rows) in downloadable, structured datasets on each platform, measuring the actual size of datasets provided. Figure 2 [Figure 2: see original paper] shows the top 10 platforms by data capacity. The gap in data capacity among platforms is significant: the top three platforms all have data capacity exceeding 80 million, while over one-third of platforms have data capacity below 100,000.

## 4.2 Data Quality

**4.2.1 High-Quality Data** High-quality data refers to datasets with large data volume and high social demand. This study sorted all downloadable datasets on each platform by data capacity and identified 146 high-quality datasets, with their distribution shown in Figure 3 [Figure 3: see original paper]. Among the 46 government open data platforms, 17 platforms had high-quality datasets selected, while no high-quality datasets were found on other platforms. Table 5 lists the top 10 high-quality datasets by name and affiliated platform. These datasets generally have high numbers of records, fields, and downloads, with content primarily related to commercial entities and pharmaceuticals.

### 4.2.2 No Low-Quality Data

- (1) **No Low-Capacity Data.** Low-capacity data refers to datasets with two or fewer rows, either due to scarce data volume or excessive granularity after statistical aggregation. Such data has low reuse value. Nearly two-thirds of the 46 open data platforms contain low-capacity data.
- (2) **No Fragmented Data.** Fragmented data refers to datasets artificially divided by time, administrative region, or government department. Integrating these datasets would better facilitate social development and utilization. Currently, most local open data platforms have fragmented data.
- (3) **No Forced Format Conversion.** Forced format conversion refers to platforms converting data from unstructured DOC, PDF, and other files into machine-readable formats like XLS or CSV through crude methods while the data remains essentially unstructured. For example, pasting large paragraphs of text from WORD files into XLS files or directly converting DOC formats to XLS formats. This study found eight local platforms with such problems.

### 4.2.3 No Problematic Data

- (1) No Duplicate Creation. Duplicate creation refers to datasets with identical titles, downloadable data files, and themes appearing repeatedly on a platform. About one-third of the 46 open data platforms have duplicate creation issues.
- (2) No Invalid Data. Invalid data refers to three situations: (a) No data available for access under the dataset name; (b) Only data links provided without accessible datasets; (c) Datasets that after downloading contain no actual data. Over half of the 46 government open data platforms contain invalid data.

## 4.3 Data Standards

**4.3.1 Open Licensing** Open data should legally guarantee openness through data open licensing agreements. Currently, platform data open licensing is typically included in website statements, disclaimer terms, or service agreements. This study found that 33 of the 46 government open data platforms have data open licensing agreements. Licensing content should explicitly grant users four rights: free access, non-discriminatory use, free utilization, and free dissemination and sharing. Only five local platforms—Beijing, Shanghai, Guizhou, Guangzhou, and Guiyang—explicitly grant all four rights in their licensing agreements. Most regions satisfy free access and non-discrimination, while others do not clearly mention or ambiguously address all four indicators.

- (1) Free Access. Free access means platforms explicitly grant users the right to freely obtain and utilize open data in licensing agreements. As public resources, government data should in principle be opened to society free of charge, except when additional value-added processing or personalized processing for minority users is required. Currently, corresponding terms on various platforms fall into two categories: “free without time limit” and “free at current stage.” Only Guizhou, Guiyang, Dongguan, and a few other regions explicitly state that data is free and without time limit, allowing users to permanently and 无偿ly access all data resources provided by the platform. Most other local platforms mention “free at current stage” in their service agreements but include vague time limits or restrictions, such as “reserving the right to charge.”
- (2) Non-Discrimination. Open licensing means platforms explicitly grant any user equal rights to access, obtain, use, and share open data. Currently, all local platforms explicitly guarantee non-discrimination in data openness, providing equal data access and utilization permissions to all users, with expressions such as “users enjoy non-exclusive use rights of data resources” and “without discrimination.”
- (3) Free Utilization. Open licensing should explicitly grant users unrestricted rights to commercially and non-commercially utilize “open data.” Currently, corresponding terms on various platforms fall into two

categories: “explicitly permits free utilization” and “does not mention free utilization.” Only Beijing, Guangzhou, Guiyang, Guizhou, and Shanghai explicitly state that users can “unrestrictedly conduct commercial and non-commercial utilization,” “enjoy rights to value-added utilization,” or “freely utilize.” Other regions’ terms do not explicitly address users’ rights to utilize data.

- (4) Free Dissemination and Sharing. Open licensing should explicitly grant users the right to freely disseminate and share open data. Currently, corresponding terms on various platforms fall into three categories: “freely disseminable,” “does not mention free dissemination,” and “free dissemination restricted.” Beijing, Guangzhou, Guiyang, Guizhou, and Shanghai grant users the right to freely disseminate existing open data.

**4.3.2 Technical Openness** Based on Tim Berners-Lee’s five-star standard for open data and other relevant open data format standards, this study assesses the format standards of datasets on local government open data platforms.

- (1) Machine-Readable Format. To facilitate user access and utilization, downloadable datasets should be in machine-readable formats such as XLS, CSV, JSON, and XML. Figure 4 [Figure 4: see original paper] compares the total number of downloadable datasets with machine-readable datasets across platforms (top 10). Overall, 38 platforms have datasets that basically meet machine-readable format requirements, though some platforms still provide datasets in non-machine-readable formats like DOC, PDF, and JPG.
- (2) Open Format. Downloadable datasets should be provided in open, non-proprietary formats, ensuring no entity can exclude others from using data through format restrictions, and that data can be accessed without requiring specific (particularly paid) software or applications. For example, CSV is an open format while XLS is not. Figure 5 [Figure 5: see original paper] compares the total number of downloadable datasets with open-format datasets across platforms (top 10). Currently, 24 local platforms provide datasets entirely in open formats, while other platforms provide no open-format datasets.
- (3) RDF Format. This study also assessed RDF format, representing the four-star requirement in the open data five-star standard. Currently, only Guiyang provides RDF-formatted datasets, with a total of 216 RDF datasets.
- (4) API Interface Proportion. In addition to direct download access, data can be provided through interfaces to enable users to efficiently obtain data in real time for application development, particularly suitable for opening large-scale, real-time data. Currently, 16 local platforms provide interfaces for each dataset, while nearly one-third of platforms provide no or only a few API interfaces.

**4.3.3 Metadata Completeness** Providing metadata helps data users clearly understand dataset content and background, enabling better access and utilization.

- (1) **Basic Metadata Coverage.** Based on China’s “Guidelines for Government Information Resource Cataloging (Trial)” regarding core metadata definitions, basic metadata entries provided on international open data platforms, and metadata entries currently provided by over half of China’s platforms, this study identified 12 entries as basic metadata for open datasets: data name, abstract/description, tags/keywords, data theme, data format, openness attribute, providing unit, release date, update date, update frequency, data indicators, and data volume. Figure 6 [Figure 6: see original paper] shows the distribution of these 12 basic metadata entries across the 46 local platforms. Currently, all 46 platforms provide dataset names, most provide abstracts, providing units, release dates, data themes, and data formats, while relatively fewer provide update frequency, data volume, and data indicators.
- (2) **API Description Standardization.** API descriptions help data users clearly understand specific API information and access methods to better call interfaces and obtain data. This study assessed API descriptions from two aspects: resource description and data call instructions. Resource description refers to basic API information such as name, description, providing department, and update time. Data call instructions refer to API call methods and request addresses. Among the 36 platforms providing API interfaces, 33 provide resource descriptions and data call instructions.

## 4.4 Data Coverage

**4.4.1 Thematic Coverage** Increasing the breadth and coverage of data openness facilitates data users’ integration and utilization of data from multiple domains. This study categorized open data themes into 14 major categories: economy and trade, transportation, institutions and organizations, culture and leisure, health and wellness, education and technology, social welfare, resources and environment, urban construction and housing, public safety, agriculture and rural areas, social security and employment, finance and taxation, and credit services. Figure 7 [Figure 7: see original paper] shows the number of datasets opened under each of the 14 themes across platforms. Among them, social welfare, economy and trade, and education and technology have the most datasets. Thematic coverage varies significantly across platforms, with Guangzhou and Qingdao covering all 14 themes.

**4.4.2 Departmental Coverage** Departmental coverage of open datasets reflects the degree of participation in data openness across government departments and the comprehensiveness of dataset sources. This study identified departments with numerous open datasets as “major data-providing departments.” Figure 8 [Figure 8: see original paper] shows the total number of datasets opened

by different types of government departments across platforms, with statistical departments opening the most data, followed by education, transportation, and health departments. However, statistical departments mostly provide macro-level data that has been processed and aggregated, with larger granularity and lower data capacity, which is less conducive to data utilization and value creation. Further analysis reveals that in places where statistical bureau data accounts for a high proportion, open data capacity is generally low (see Figure 9 [Figure 9: see original paper]).

**4.4.3 High-Demand Keyword Coverage** This study conducted text analysis on dataset names from the top 20% most downloaded datasets available on each platform, identifying frequently occurring keywords that reflect the content and distribution of high-demand open datasets. Figure 10 [Figure 10: see original paper] shows the most frequently occurring descriptive qualifiers, such as “enterprise,” “license,” “construction,” and “production.”

#### 4.4.4 Data Sustainability

- (1) Continuous Growth. This study judged whether platform datasets achieved continuous growth based on dataset creation dates, tracking and analyzing by quarter. Figure 11 [Figure 11: see original paper] shows the continuous growth of datasets on national platforms by quarter since platform launch, with colored areas indicating periods with new datasets and blank areas indicating periods without new datasets. The Shanghai platform has maintained continuous growth for 15 quarters since launching data in Q4 2012. New platforms launched since 2016 have mostly maintained regular dataset growth.
- (2) Dynamic Updates. This study tracked the number of datasets updated on local data platforms from January to April 2018, as shown in Figure 12 [Figure 12: see original paper]. The Guiyang platform updated the most datasets during this period, exceeding 1,000, while a few local platforms had no data updates at all.
- (3) Historical Archiving. Historical archiving means platforms retain multiple historical batches of data simultaneously for user download, facilitating time-series data access and utilization. Figure 13 [Figure 13: see original paper] shows the distribution of platforms implementing data historical archiving. Currently, 15 local platforms including Shanghai, Guangdong, and Guangzhou have achieved data historical archiving.

## 5 Research Conclusions

### 5.1 From National to Local Level

Currently, several authoritative international assessment reports on government data openness have been released, with the two most influential being the “Open

Data Barometer” published by the Web Foundation and the “Global Open Data Index” published by the Open Knowledge Foundation. The Open Data Barometer, launched by the Web Foundation in 2013, combines expert surveys, auxiliary data, peer assessment, quantitative data, and qualitative assessment to evaluate government data openness in approximately 115 countries every two years across three layers: “readiness,” “implementation,” and “impact.” In 2016, China ranked 71st. The Global Open Data Index, conducted by the Open Knowledge Foundation, primarily assesses key datasets opened by countries using snowball sampling, volunteer questionnaires, interviews, continuous crowdsourcing, expert assessment, and peer assessment. In 2015, the index assessed 122 countries and regions, with China ranking 93rd.

International assessment reports reveal that China’s government data openness still lags globally and requires improvement. Moreover, international assessment systems and methods are not entirely applicable to China’s actual conditions and development stage. For example, assessment indicators for legal policies and institutional mechanisms cannot be directly applied to the Chinese context, and assessments of utilization effects and impacts post-openness are somewhat premature for China’s current data openness status. Furthermore, these assessment reports have not evaluated local-level data openness in China.

Therefore, this study constructs a more China-appropriate assessment framework targeting the previously unaddressed local government level of data openness and conducts actual evaluation. The framework focuses on the core of current Chinese local government data openness—the data dimension—while incorporating actual development stages and social needs, emphasizing assessment dimensions including data quality, data standards, data sustainability, data quantity, and data coverage. Compared with existing international assessment systems, this framework better promotes the development of Chinese local government data openness. Additionally, compared with previous domestic assessments, this study’s sample is more comprehensive, covering 46 provincial and prefecture-level governments in China, and the assessment indicators are more focused.

## 5.2 Steady Growth in Data Quantity but Low Dataset Capacity

Research finds that the total number of open datasets in China is growing steadily, with Guiyang, Shanghai, Qingdao, Wuhan, and other locations exceeding 1,000 datasets, while over 20 other locations still have fewer than 200 open datasets. More importantly, the overall data capacity of open datasets is low, with most local datasets having insufficient rows and columns, making them difficult to utilize effectively.

## 5.3 Uneven Data Quality with Widespread Problematic Data

Some regions have launched batches of high-capacity, high-demand, high-quality datasets, but most local datasets are of uneven quality, with widespread low-

quality data issues including low-capacity fragmented data, duplicate creation, forced format conversion, and invalid data. These low-quality and problematic datasets are difficult to reuse and generate value, reducing data openness to a mere formality.

#### **5.4 Data Standards Require Regulation and Improvement**

In terms of legal openness, while a few local platforms explicitly grant users four rights—free access, non-discrimination, free utilization, and free dissemination and sharing—in their licensing agreements, many local government data platforms still lack clear and adequate data open licensing.

In terms of technical openness, the proportion of machine-readable, non-proprietary, interface-provided datasets on various platforms is growing steadily. Guiyang has pioneered the provision of RDF-formatted datasets in China. However, many datasets nationwide still do not comply with open data format standards.

Regarding metadata completeness, while most local platforms can provide basic metadata, the situation varies across regions, with widespread lack of dataset release time, update time, data indicators, and data volume entries.

#### **5.5 Low Data Coverage Dominated by Statistical Data**

In terms of thematic coverage, the most provided dataset themes on current platforms are social welfare and economy and trade, but thematic coverage varies significantly across platforms.

In terms of departmental coverage, statistical departments open the most datasets on various platforms. However, statistical departments mostly release secondary, processed, and aggregated datasets with lower reuse value compared to primary, original datasets from business departments. Nearly half of local governments still have less than half of their departments participating in data openness.

In terms of keyword coverage, high-demand open dataset names most frequently include “enterprise,” “license,” “construction,” and “production,” though a few cities have coverage rates below 30%.

#### **5.6 Low Data Sustainability with Insufficient Daily Updates and Growth**

Only a few localities can basically maintain continuous growth of new datasets and dynamic updates of existing datasets, but many platforms experience intermittent or stagnant dataset growth, with the proportion truly achieving dynamic updates of existing data remaining low. Additionally, fewer than half of localities retain multiple historical batches of data on platforms for user access. Most local platforms do not pay sufficient attention to continuous operation and updates of open data.

## 6 Recommendations

### 6.1 Increase Data Quantity

Both the total number of open datasets and dataset capacity—that is, the number of fields (columns) and records (rows) in open datasets—should be increased.

### 6.2 Improve Data Quality

Open high-value, high-social-demand quality datasets. Casually releasing easily published, low-density, fragmented, problematic data does not create value. Open data should start from actual user needs rather than government departments' self-judgment, regularly soliciting needs and suggestions from data users, and targeted opening of datasets that society truly needs and that can solve problems and create value, while ensuring data completeness, accuracy, and applicability.

### 6.3 Standardize Data Standards

**6.3.1 Legal Openness** Provide licensing agreements for open data that explicitly grant users rights to free access, non-discriminatory use, free utilization, and free dissemination and sharing. Further explore graded and classified approaches to equip different datasets with different licensing content.

**6.3.2 Technical Openness** Based on basic principles and standards of open data, open complete, original, machine-readable, open-format, structured, and electronic datasets that users can actually use. According to Tim Berners-Lee's five-star standard, most Chinese open government data currently meets three-star standards. The next step for local government data openness should advance toward four-star RDF format and continue developing toward five-star standards to enable data interlinking.

In addition to ensuring direct download access, government open data platforms should also open large-scale, real-time, high-processing-requirement data through API interfaces. API interfaces should also be provided with standardized descriptions, including resource descriptions and data call instructions, to help data users understand specific API information and access methods for better interface calling and data acquisition.

**6.3.3 Metadata Completeness** Platforms should provide comprehensive metadata information when opening datasets to help data users clearly understand dataset content and background for better comprehension and utilization. Metadata entries can include data name, abstract/description, tags/keywords, data theme, data format, openness attribute, providing unit, release date, update date, update frequency, data volume, and field names.

#### 6.4 Expand Data Coverage

Maximize coverage of key datasets in priority openness areas to improve data breadth, richness, and relevance, enabling data users to fully access and integrate multi-source data for deep mining and utilization. Governments should focus on improving participation in data openness across business departments rather than concentrating on statistical departments.

#### 6.5 Maintain Data Sustainability

Open government data is a continuous and normalized task requiring ongoing updates and increases after datasets are opened. Only continuous data supply can stimulate data utilization vitality and meet society's growing demand for open data. Governments should establish long-term working mechanisms to ensure dynamic updates of existing open datasets, continuous addition of new datasets, and retention of historically opened data batches on platforms for continued user access and utilization.

### References

- [1] Zheng Lei. Value creation mechanism of open government data: An ecosystem perspective [J]. *E-Government*, 2015(7): 2-7.
- [2] The annotated 8 principles of open government data. Open government data principles [EB/OL]. [2018-05-04]. [https://public.resource.org/8\\_{principles}.html](https://public.resource.org/8_{principles}.html).
- [3] Open definition. The open definition [EB/OL]. [2018-05-04]. <https://opendefinition.org>.
- [4] The World Bank. Open data essentials [EB/OL]. [2018-05-04]. <http://opendatatoolkit.worldbank.org/en/essentials.html>.
- [5] Linked Data [EB/OL]. [2017-08-15]. <https://www.w3.org/DesignIssues/LinkedData.html>.
- [6] Open Data Charter. Principles [EB/OL]. [2018-05-04]. <https://opendatacharter.net/principles>.
- [7] OVIENDO E, MAZON J N, ZUBCOFF J J. Towards a data quality model for open data portals [C]//Computing Conference (CLEI), 2013 XXXIX Latin American computing conference. Naiguata: IEEE, 2013: 1-8.
- [8] VISCUSI G, SPAHIU B, MAURINO A, et al. Compliance with open government data policies: An empirical assessment of Italian Local Public Administrations [J]. *Information policy*, 2014, 19(3/4): 263-275.
- [9] LOURENÇO R P. An analysis of open government portals: A perspective of transparency for accountability [J]. *Government information quarterly*, 2015, 32(3): 323-332.
- [10] BELLO O, AKINWANDE V, JOLAYEMI O, et al. Open data portals in Africa: An analysis of open government data initiatives [J]. *African journal of library, archives & information science*, 2016, 26(2): 97-106.
- [11] The GovLab. Open data definitions [EB/OL]. [2017-08-15]. <http://odimpact.org/resources.html>.
- [12] SUSHI I, ZUIDERWIJK A, JANSSEN M, et al. Benchmarks for evaluating the progress of open data adoption: Usage, limitations, and lessons learned [J]. *Social science computer review*, 2015, 33(5): 613-630.
- [13] Xia Yijian. Comparison and analysis of open government data assessment methods by international organizations [J]. *Library and Information Service*, 2015, 59(19): 75-83.
- [14] Zheng Lei, Guan Wenwen. Research on open government data

assessment systems, indicators, and methods [J]. *Library and Information Service*, 2016, 60(18): 43-55. [15] Zheng Yueping, Liu Meicen. Current status and existing problems of open data assessment—Based on comparison and analysis of foreign open data assessments [J]. *E-Government*, 2016(8): 84-93. [16] Chen Mei. Value assessment of open government data: Progress and implications [J]. *Journal of Intelligence*, 2017(11): 92-98. [17] Wei Xinling, An Xiaomi, Li Xuemei, et al. Review of open government data assessment systems: Characteristic analysis [J]. *Library and Information Service*, 2017, 61(18): 119-127. [18] Zheng Lei, Gao Feng. Research on China's open government data platforms: Framework, status, and recommendations [J]. *E-Government*, 2015(7): 8-16. [19] Zheng Lei, Xiong Jiuyang. Research on Chinese local government open data: Technical and legal characteristics [J]. *Public Administration Review*, 2017(1): 53-74. [20] Xia Yijian. Current status, problems, and countermeasures of China's open government data from an international comparative perspective [J]. *Library and Information Service*, 2016, 60(7): 34-40. [21] Zhao Jidi, Zhang Hanlun. Research on evaluation of local government data openness effectiveness—A case study of Shanghai [J]. *E-Government*, 2017(9): 11-21. [22] Shen Jing, Han Lei, Hu Guangwei. Research on government data openness development speed index—Based on assessment of China's provincial government open data platforms [J]. *Journal of Intelligence*, 2018(8): 1. [23] Hailun, Deng Song. Efficiency assessment of city government open data platforms based on data envelopment analysis [J]. *E-Government*, 2018(8): 112-118.

## Author Contributions

**Zheng Lei:** Responsible for literature review, determining the assessment framework and methodology, and writing the analysis.

**Lü Wenzeng:** Responsible for crawling and analyzing data and producing charts.

---

**Zheng Lei, Lü Wenzeng. Assessing Open Data at Local Government Level: Framework and Findings.** School of International Relations and Public Affairs, Fudan University, Shanghai 200433.

**Abstract:** [Purpose/significance] This paper attempts to construct an assessment framework on open government data at local government level, evaluate the data dimension on existing local government open data platforms in China and put forward suggestions to foster the opening of local government data. [Method/process] Based on the definitions, principles and standards of open data, learning from international open data assessment frameworks, taking into consideration of the policy requirement and development status of open data practices in China, this paper constructs a systematic, multi-dimensional and operable assessment framework, and carries out an actual assessment on forty-six local government open data platforms in China. [Result/conclusion] The

study finds out a number of problems with regard to the quantity, quality, standard, coverage and sustainability of open data on local government platforms in China.

**Keywords:** China; local government; open data; assessment; framework

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

*Source: ChinaXiv — Machine translation. Verify with original.*