Comparative Analysis of Metadata Schemes for China-US Open Government Data Platforms (Postprint)

Authors: Yang Sinan

Date: 2023-10-08T00:00:00+00:00

Abstract

[Purpose / Significance] Open government data facilitates the rational development and utilization of data resources, encourages social innovation, and promotes economic development. Ensuring the effective utilization and social value-added of open government data relies on the support of high-quality metadata schemes. [Method / Process] Through research and analysis of relevant domestic and international research findings on open government data, this study investigates the metadata schemes of major local government open data platforms in China and conducts a comparative analysis with the U.S. Government's open government data metadata standards. [Results / Conclusion] The results indicate that China's local government open data suffers from deficiencies such as inconsistent metadata schemes, simple and incomplete descriptions of datasets, formats primarily based on HTML web pages, and low machine readability, which affect the effectiveness of open data utilization. Therefore, the Chinese government should draw upon mature and effective international metadata scheme standards, formulate unified and standardized metadata schemes, and ensure the provision of high-quality, high-value data that meets societal needs.

Full Text

Preamble

A Comparative Analysis of Metadata Schemes for Chinese and American Open Government Data Platforms

School of Information Management, Sun Yat-sen University, Guangzhou 510006

Abstract

[**Purpose/Significance**] Open government data facilitates the rational development and utilization of data resources, encourages social innovation, and promotes economic development. Ensuring the effective use and social value-added utilization of open government data relies on the support of high-quality metadata schemes. [**Method/Process**] This study analyzes relevant research on open government data both domestically and internationally, investigates the metadata schemes of major local government open data platforms in China, and conducts a comparative analysis with the metadata standards of the U.S. government open data platform. [**Results/Conclusion**] The findings reveal that Chinese local government open data platforms suffer from inconsistent metadata schemes, simplistic and incomplete dataset descriptions, predominant use of HTML web formats, and low machine readability, all of which affect the usability of open data. Therefore, the Chinese government should draw upon mature and effective international metadata standards to formulate unified and standardized metadata schemes, ensuring the provision of high-quality, high-value data that meets societal needs.

Keywords: open government data; open data platform; metadata scheme

1 Introduction

Open government data refers to the proactive release of government-owned data that does not involve personal privacy or public security, made freely available to all citizens through modern information technology. Establishing official platforms for government data management, release, query, and reuse, and publishing open data catalogs constitute key initiatives universally adopted in the global open government data movement.

Since the launch of the U.S. government data platform Data.gov in 2009, the open data movement has rapidly gained momentum worldwide. Countries including the United Kingdom, Australia, and Canada have successively established government open data platforms. By 2015, 92 countries had been included in the global "Open Data Barometer" assessment report published by the World Wide Web Foundation, founded by Tim Berners-Lee, the inventor of the World Wide Web.

China's research and application of open government data is still in its infancy. As early as 2011, cities such as Shanghai and Beijing pioneered experiments in government data openness, achieving certain results. However, at the national level, China has not yet established a unified website or platform for open government data, making it impossible to integrate government data vertically and horizontally, with data remaining fragmented. Since 2014, China has begun discussing relevant plans and proposals for open data. In September 2015, the State Council issued the "Outline for Promoting the Development of Big Data," explicitly proposing the establishment of a unified national government data open platform by the end of 2018 [1].

Academic discussions and research on open government data in China began with the U.S. government's launch of the open data movement in 2009. Through literature review, current domestic research primarily focuses on three aspects: (1) Conceptual analysis and definition of open data. Scholars such as Tan Jian [2], Li Jiajia [3], Xiang Liling et al. [4], and Zheng Lei [5] have proposed perspectives from different angles. (2) Theoretical foundation building and future prospects. Some scholars have explored the connotation, theoretical basis, and policy foundation of open government data, discussing its social significance and importance from perspectives of e-government, public management, and legal construction, such as Tan Jian [2], Xiang Liling et al. [4], and Ma Haiqun et al. [6]. (3) Summary of domestic practices and foreign experience. Research and practice in countries like the UK and the US preceded China, and scholars such as Lu Jianying et al. [7], Ma Haiqun et al. [8], Wu Min [9], Chen Mei [10-11], and Cao Ling [12] have studied and discussed foreign practices. Additionally, scholars including Zheng Lei et al. [13], Huang Simian et al. [14], Zhou Wenhong [15], and Huang Ruhua et al. [16] have investigated and summarized China's practical achievements in recent years.

Metadata is data about data and is itself structured data. Introducing information organization into website construction enables effective organization of web information resources [17]. Using metadata to describe data information resources in website information organization not only facilitates user understanding but also enables easier data processing by machines or computer programs, thereby enhancing resource operability. Therefore, ensuring the use value of open data relies on scientific metadata scheme support; the higher the data quality and the stronger the descriptive power of metadata, the greater the opportunities for data discovery and utilization. Research on metadata schemes for open government data platforms has been conducted by a few domestic scholars in recent years. For instance, Yu Mengyue et al. [18] investigated the metadata schemes of government open data platforms in Beijing and Zhejiang, proposing that adopting internationally 通用的 metadata standards could improve metadata quality. Wu Gang et al. [19] selected government open data platforms from the US, UK, Australia, Canada, as well as Beijing and Shanghai in China as research objects, comparing resource status, organization, and retrieval aspects, and proposing construction strategies. Zhao Rongying et al. [20] examined the UK government open data website from the perspective of data sharing, summarizing metadata standard characteristics from document structure, element composition, and rules.

Domestic research on metadata scheme standards for open government data remains limited, whereas internationally, numerous studies exist on metadata standardization. For example, C. Turbelin et al. [21] pointed out through investigation that the lack of unified standards in open information platforms leads to limited utilization of data resources. Scholars and organizations have also proposed data standards, including the 8 Principles of Open Government Data [22], official guidelines published by the W3C eGov Interest Group [23], and technical standards for linked open data [24].

China's open government data platform construction, particularly regarding metadata schemes, still has considerable room for development and improvement to achieve standardization, interoperability, and machine readability, which is crucial for enhancing the usability of open data resources, improving citizens' lives, facilitating scientific research, and building smart cities. Therefore, this study analyzes relevant research on open government data both domestically and internationally, selecting the U.S. open government data platform data.gov and major Chinese local government open data platforms for comparative analysis of metadata schemes, aiming to provide reference suggestions for constructing a unified metadata scheme standard for China's open government data platform.

2 Metadata Schemes for Foreign Open Government Data Platforms

2.1 International Metadata Standards for Open Government Data

On January 16, 2014, the Government Linked Data Working Group officially released the Data Catalog Vocabulary (DCAT) [25], drafted by the Digital Enterprise Research Institute (DERI) at the National University of Ireland and revised by the W3C eGovernment Interest Group.

The DCAT vocabulary includes 7 classes and 17 properties, reusing relevant terms from Dublin Core, FOAF ontology, and SKOS ontology, and supporting interoperability between data catalogs to compensate for resource sharing deficiencies caused by diverse metadata applications, thereby satisfying user needs for accessing various information resources through a unified interface. The main classes and their properties are shown in Figure 1 [Figure 1: see original paper] [25].

As summarized from Figure 1, the DCAT vocabulary comprehensively covers data attributes, providing fundamental properties to help users understand basic dataset information, such as dct:title, dcat:theme, and dcat:keyword; properties to help users filter data, such as dct:description; and properties to help users locate data, such as dct:modified.

According to W3C statistics, DCAT and its Application Profile have been adopted by governments, organizations, and institutions in multiple countries, including the European Union, for open data practices [26], demonstrating DCAT's broad applicability and scientific validity.

2.2 Metadata Scheme of the U.S. Open Government Data Platform

From a global perspective, the U.S. is a pioneer in the open data movement and leads in open government data. The U.S. ranked in the top ten of the "Open Data Barometer" assessment for three consecutive years from 2013 to 2015 [27]. U.S. governments at all levels attach great importance to the standardization of metadata for open government data. The national data portal's metadata

scheme has established mappings to both DCAT and Schema.org, with field descriptions provided in the schema [28]. By August 2016, the number of local government open data portals in the U.S. had reached 88. For example, the New York State government launched its open data platform Open.ny.gov in March 2013 and subsequently published the "Open Data Handbook," which detailed the construction principles and schemes for metadata. Therefore, selecting the U.S. as a reference object holds significance for Chinese local governments to develop and improve their open government data platform metadata schemes.

Through investigation of the U.S. open government data platform data.gov, it can be observed that each dataset detail page includes a dedicated metadata description section, as shown in Figure 2 [Figure 2: see original paper]. In addition to displaying and explaining metadata elements in HTML format on the webpage, downloadable metadata files use JSON syntax to normalize the expression of multi-valued, multi-group metadata attributes with high machine readability, allowing users to download corresponding metadata JSON files locally.

Taking the U.S. Housing Affordability Data System dataset (https://catalog.data.gov/dataset/housing-affordability-data-system-hads) as an example, the dataset's metadata description elements are illustrated in Figure 3 [Figure 3: see original paper]. This dataset contains 14 top-level metadata elements, among which contactPoint, distribution, and publisher each include several second-level fields. The U.S. government open data platform metadata scheme demonstrates high interoperability, incorporating multiple data catalogs including DCAT, vCard (for describing personal business card information), and Org (for describing organizational information).

The U.S. government released the updated Project Open Data Metadata Schema v1.1 on February 1, 2015 [28], which lists the composition of dataset metadata scheme elements as shown in Figure 4 [Figure 4: see original paper].

Based on the actual datasets on the U.S. government open data platform data.gov, this study summarizes the detailed information of the main elements of the U.S. government open data platform metadata scheme, as presented in Table 1 .

As shown in Table 1, the metadata scheme adopted by the U.S. government open data platform comprehensively describes datasets, including not only descriptive metadata such as title, date, description, and keywords, but also data information such as temporal and spatial applicability. Furthermore, it provides standardized digital symbol identifiers for normative description of attribute content whenever possible, such as dataset identifiers and administrative system codes for publishers, which greatly enhances the standardization of data resource description, facilitates user search and retrieval, reduces data preprocessing work, and improves user efficiency.

Overall, the metadata scheme of the U.S. government open data platform helps users understand data information and quickly grasp data content, offers high

operability for machine processing and analysis, and addresses resource interoperability issues through mappings between multiple metadata scheme standards.

3 Metadata Schemes for Chinese Open Government Data Platforms

3.1 Current Status of Metadata Schemes on Chinese Local Government Platforms

The construction of open government data platforms constitutes an important component of government data openness. While the open data movements in the U.S. and UK developed from central to local governments, China's government open data platform construction has been driven by autonomous practices at local levels under central policy guidance. According to statistics, as of April 2017, 19 prefecture-level and above local governments in China had established open data platforms, as detailed in Table 2 .

Investigation of these local government open data platforms reveals that although basic metadata information is provided, the metadata schemes vary across platforms, predominantly using HTML web formats with simple page descriptions that cannot comprehensively describe datasets. None of the platforms provide a unified metadata scheme introduction report. Only the Guangzhou Municipal Government Data Unified Open Platform provides textual descriptions of metadata for each dataset. For example, under the dataset "Guangzhou Scientific and Technological Achievement Registration Information," metadata includes year, enterprise name (achievement completion unit name or completer), address, contact phone, achievement name, registration time, and update time, as shown in Figure 5 [Figure 5: see original paper]. The Guangzhou platform uses dataset fields as metadata elements rather than a consistent metadata scheme. While adding annotations and tags for specific datasets facilitates personalized metadata utilization, it leads to low standardization and normalization of metadata element settings and significant differences in metadata composition between datasets.

Moreover, Chinese local government open data platforms do not provide metadata files, displaying metadata only on webpages, which affects the effectiveness of open data utilization and hinders the achievement of standardization, normalization, interoperability, and machine readability. By examining HTML page element tags, this study summarizes the core metadata element composition of some local government open data platforms, as shown in Table 3 .

As evident from Table 3, current metadata settings across local government open data platforms primarily include: basic descriptive metadata such as title, abstract, keywords, and publisher; basic administrative metadata such as resource creation time and update time; and a small amount of usage metadata such as user access and download statistics. The Shanghai Municipal Government Data Service Network (http://www.datashanghai.gov.cn) started early and has devel-

oped rapidly, outperforming other provinces and municipalities in terms of data volume, openness, timeliness, and usability. Compared with other Chinese local government platforms, Shanghai's metadata scheme is relatively comprehensive and detailed.

Furthermore, Table 3 shows that Chinese local government open data platforms' metadata schemes basically reference and follow core elements from government information disclosure metadata schemes (see Table 4 [29]), with connotations largely consistent with Dublin Core elements, indicating that Chinese open government data platform metadata schemes remain at the stage of government information disclosure.

3.2 Deficiencies and Recommendations for Chinese Local Government Platforms

As information centers aggregating large volumes of data from multiple government agencies and departments, open government data platforms serve not only conventional functions of information collection, organization, publication, and dissemination, but more importantly, provide public government datasets to all sectors of society. The sharing and utilization of open government data depends on metadata support, and attaching high-quality metadata descriptions to datasets forms the foundation for interoperability between data catalogs. Although current metadata settings appear to meet basic needs for local government open data, they lag far behind international standards in terms of user query, usage, and machine readability. Through comparative analysis with the U.S. government open data platform metadata scheme, Chinese local government open data platform metadata schemes exhibit several major deficiencies:

(1) **Lack of unified standards across local governments.** Since China's government data openness movement has been initiated autonomously by local governments, significant variations in metadata schemes have emerged, primarily manifested in inconsistent element composition, naming, and formats. This creates difficulties for application development based on government open data, as the lack of unified dataset metadata schemes results in high processing complexity and costs for reading data across multiple catalogs.

(2) **Absence of machine-readable metadata formats.** Local government open data platforms only display metadata on HTML pages without providing downloadable metadata files, which is detrimental to data interoperability and machine processing. Currently, the widely used international metadata file formats are JSON and XML.

(3) **Incomplete metadata element composition.** Existing local government open data platform metadata lacks essential elements such as unique resource identifiers, which compromises data standardization and stability. They also fail to provide common metadata elements for resource references and linkages, open licensing, and spatial-temporal coverage.

(4) **Lack of detailed platform metadata scheme documentation.** The absence of metadata scheme manuals and usage instructions increases user difficulty and costs, reducing motivation and enthusiasm for developing government open data applications and hindering value-added data utilization.

(5) **Inconsistent metadata formats within the same platform.** For example, the Zhejiang Provincial Government Open Data Platform exhibits inconsistent value formats for "file size" across different datasets, with some including units (e.g., 100KB) and others omitting them (e.g., 100).

Based on these deficiencies, the Chinese government should actively standardize platform resource organization and description methods, draw upon mature and effective international metadata scheme standards, and formulate unified and standardized metadata schemes that comprehensively describe data resources and facilitate user search, retrieval, browsing, and identification. Strict standardization of value formats for each metadata field and provision of detailed metadata scheme reports will enhance data openness and machine readability, ensuring platforms can truly provide high-quality, high-value data that meets societal needs.

4 Conclusion

Open government data represents a global wave, reflecting new public demands for government management approaches and the needs of national economic development. The trend of Chinese government open data is irreversible, though significant gaps remain compared to developed countries like the U.S. Existing Chinese local government open data platforms suffer from low resource conversion rates. By improving the scientific organization of information resources and adopting unified metadata standards for data description, platforms can more efficiently and rapidly meet user needs from institutions, organizations, and individuals during query, retrieval, identification, storage, transmission, utilization, and secondary development processes, truly achieving the goals of information benefitting the public and promoting social development through open data. In terms of technical implementation for open government data, China can directly adopt internationally advanced approaches that best facilitate data value realization to rapidly enhance data application capabilities. Regarding the standardization and normalization of metadata for open government data, drawing upon and referencing internationally 通用 metadata standards will be an important development direction.

References

[1] State Council Notice on Issuing the Outline for Promoting the Development of Big Data [EB/OL]. [2017-06-15]. http://www.gov.cn/zhengce/content/2015-09/05/content_{10137}.htm.
[2] Tan J. Open Data and Its Application Status [J]. Library and Information,

2011(4): 50-54.
[3] Li J J. A New Perspective on Information Management: Open Data [J]. Information Theory and Practice, 2010, 33(7): 35-39.
[4] Xiang L L, Wang Q. On the Legal Attributes, Obligations and Related Mechanisms of Open Data [J]. Journal of the National Library of China, 2013, 22(5): 38-44.
[5] Zheng L. Research on Open Government Data: Conceptual Analysis, Key Factors and Their Interactions [J]. Chinese Public Administration, 2015(11): 13-18.
[6] Ma H Q, Pu P. Analysis of the Connotation and Theoretical Basis of Open Data [J]. Library Theory and Practice, 2016(11): 48-54.
[7] Lu J Y, Zheng L, Sharon S D. U.S. Government Data Openness: History, Progress and Implications [J]. E-Government, 2013, 16(6): 26-32.
[8] Ma H Q, Pu P. 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): 76-87.
[9] Wu M. Application and Implications of Open Data in UK and US Governments [J]. Library and Information, 2012(1): 127-130.
[10] Chen M. Research on the Guarantee Mechanism of U.S. Open Government Data [J]. Journal of Intelligence, 2013, 32(7): 148-153.
[11] Chen M. Research on the Implementation of UK Open Data Policy [J]. Library Development, 2014(3): 22-27.
[12] Cao L. Big Data Innovation: Research on EU Open Data Strategy [J]. Information Theory and Practice, 2013, 36(4): 118-122.
[13] Zheng L, Gao F. Research on China's Open Government Data Platform: Framework, Current Status and Recommendations [J]. E-Government, 2015(7): 8-16.
[14] Huang S M, Zhang Y H. Research on Problems and Countermeasures in the Construction of Current Chinese Government Data Open Platforms—Taking Beijing and Shanghai Government Data Open Websites as Examples [J]. China Management Informationization, 2015(14): 175-177.
[15] Zhou W H. Progress, Problems and Countermeasures of Provincial Open Government Data in China [J]. Library and Information Knowledge, 2017(1): 11-18.
[16] Huang R H, Wang C Y. Investigation and Analysis of the Current Status of Chinese Government Data Open Platforms [J]. Information Studies: Theory & Application, 2016, 39(7): 50-55.
[17] Jiang L T. Information Organization in Government Websites [J]. China Science and Technology Information, 2012(6): 81.
[18] Yu M Y, Zhai J, Lin Y. Research on Core Metadata for Local Government Open Data in China [J]. Journal of Intelligence, 2016, 35(12): 98-104.
[19] Wu G, Zeng L Y. Comparative Study on Domestic and Foreign Government Open Data Platform Construction [J]. Information and Documentation Services, 2016(6): 75-79.
[20] Zhao R Y, Liang Z S, Duan P P. Metadata Standards for UK Government Data Openness and Sharing—Investigation and Implications of Data.gov.uk

[J]. Library and Information Service, 2016, 60(19): 31-39.
[21] Turbelin C, Boelle P Y. Open Data in Public Health Surveillance Systems: A Case Study Using the French Sentinelles Network [J]. International Journal of Medical Informatics, 2013, 82(10): 1012-1021.
[22] Open Government Data Principles [EB/OL]. [2017-06-15]. https://public.resource.org/8_{principles}.html
[23] Martin S, Foulonneau M, Turki S. 1-5 Stars: Metadata on the Openness Level of Open Data Sets in Europe [M]//Metadata and Semantics Research. Berlin: Springer International Publishing, 2013: 234-245.
[24] Bennett D. Publishing Open Government Data [EB/OL]. [2017-06-15]. https://www.w3.org/TR/gov-data.
[25] W3C. Data Catalog Vocabulary (DCAT) [EB/OL]. [2017-06-15]. https://www.w3.org/TR/vocab-dcat.
[26] W3C. DCAT Implementations [EB/OL]. [2017-06-15]. https://www.w3.org/2011/gld/wiki/DCAT_{Imple
[27] Open Data Barometer. Global Report [EB/OL]. [2017-06-15]. http://opendatabarometer.org.
[28] Project Open Data Metadata Schema v1.1 [EB/OL]. [2017-06-15]. https://project-open-data.cio.gov/v1.1/schema.
[29] Zhang X, Wang J M. Preliminary Study on Government Information Disclosure Metadata Scheme [J]. Library Development, 2008(12): 10-14.

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