Metadata Analysis and Management of Urban Sustainable Development Capability Evaluation Indicators in China: Postprint

Authors: Dong Rencai, Wang Tao, Zhang Yonglin, Zhang Xueqi, Li Huanhuan

Date: 2018-06-09T00:00:00+00:00

Abstract

Amidst China' s vigorous promotion of urban sustainable development and the advancement of national sustainable development experimental zone construction, determining which assessment methods and data to employ for evaluating urban sustainable development capacity constitutes a critical issue requiring resolution. Metadata theory and technology, which have gained prominence in recent years, are considered an effective approach for addressing data quality control in assessment processes. Targeting the characteristics of current urban sustainable development capacity assessment indicator systems utilized in China, this study proposes specific methodologies for developing an urban sustainable development capacity assessment metadata management system from a software engineering perspective, through in-depth analysis of each indicator' s data source, acquisition methods, applicable approaches, and other attributes, thereby enabling sustainable development experimental zones to efficiently obtain and manage requisite assessment data. Drawing upon the metadata standard established in the "12th Five-Year Plan" Science and Technology Support Program project "Research and Demonstration of Key Technologies for Urban Sustainable Development Capacity Assessment and Information Management", fourteen items—including "data release date", "data release format", "spatial extent", "temporal extent (start time, end time)", "statistical frequency", "data security classification", "data lineage description", "online resource link address", and "data statistical unit information (unit name, contact person, phone number, unit address, email address)" —are designated as key metadata elements for tracking benchmark assessment data. Through comparative analysis using a quantitative data quality scoring method, evaluation results demonstrate that data quality for assessed indicators exhibits substantial improvements in reliability, comparability, and sustainability scores when supported by metadata. The research concludes that metadata theory offers distinct advantages in controlling and ensuring data quality for urban sustainable development capacity

assessment, and that developing a dedicated metadata management system can effectively enhance comprehensive evaluation outcomes.

Full Text

Metadata Analysis and Management for Urban Sustainable Development Capability Evaluation Indicators in China

DONG Rencai[1], WANG Tao[1], ZHANG Yonglin[1], ZHANG Xueqi[1], LI Huanhuan[1,2]

[1]State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China

Abstract

As China vigorously promotes urban sustainable development and advances the construction of national sustainable development experimental zones, identifying scientific assessment methods and data sources for evaluating urban sustainable development capability has become a critical issue. In recent years, metadata theory and technology have emerged as effective approaches for controlling assessment data quality. This study analyzes the characteristics of current urban sustainable development capability evaluation indicator systems in China by examining each indicator' s data source, collection method, and application characteristics. We propose a software engineering-based approach to developing a metadata management system for urban sustainable development capability evaluation to help experimental zones efficiently acquire and manage required data. Drawing from the metadata standard established under the 12th Five-Year Science and Technology Support Program project "Research and Demonstration on Key Technologies for Urban Sustainable Development Capability Assessment and Information Management," we identify fourteen key metadata items—including data release date, release format, spatial range, temporal range (start and end time), statistical frequency, data security classification, data lineage description, online resource link address, and statistical unit information—as critical for tracking evaluation data. Using a quantitative data quality scoring method, we compare data quality evaluation results before and after applying metadata tracking. The findings demonstrate that when evaluation indicators are supported by metadata, their data reliability, comparability, and sustainability scores improve significantly. This research confirms that metadata theory offers distinct advantages in controlling and ensuring data quality for urban sustainable development capability assessment, and that developing a targeted metadata management system can effectively enhance comprehensive evaluation of assessment data.

Keywords: urban sustainable development capability; metadata management system; data quality evaluation; sustainable development experimental zone

1. Assessment Data Source Analysis

Research on urban sustainable development capability assessment has primarily focused on designing and innovating evaluation indicator systems and methods. However, data quality represents a key factor influencing the scientific validity of assessment results. Urban sustainable development capability assessment exhibits considerable particularity and complexity due to varying urban development models, city sizes, and differences among evaluation systems, making data sources exceptionally complex. Analysis of data sources for the urban sustainable development capability evaluation indicator system (hereinafter referred to as the evaluation indicator system) from the Science and Technology Support Program project "Research and Demonstration on Key Technologies for Urban Sustainable Development Capability Assessment and Information Management" reveals that data originates from three primary sources: statistical yearbooks and bulletins from 12 relevant departments, accounting for 68% of all sources; research reports comprising 22%; and other sources making up the remaining 10%.

Urban sustainable development assessment involves extensive elements and rich content, with comprehensive statistical data available only through relevant statistical yearbooks and bulletins. According to China' s Statistics Law, only economic and social data recognized and released by the government possess credibility, which forms the basis for recognized urban sustainable development capability assessment results. Statistical yearbooks and related bulletins, as the most common forms of government data release, constitute the main source for evaluation data. However, current data collection methods face challenges including data accuracy issues, excessive connotation, and inconsistent collection approaches [Figure 1: see original paper].

2. Data Quality Evaluation Method

The United Nations Statistical Office' s Handbook of Statistical Organization proposes eight requirements for official statistics: user-oriented services, multi-user targeting, organic data integration, timely collection and release, historical continuity, protection of respondent interests and confidentiality, objectivity and impartiality, and administrative competence. This framework has expanded the concept of statistical data quality from narrow accuracy to broader dimensions encompassing reliability, validity, and methodological professionalism.

To visually represent data quality in urban sustainable development capability assessment, this study establishes a quantitative evaluation system guided by multidimensional data quality requirements, synthesizing evaluation results for indicator system data [15-16]. The evaluation framework addresses three dimen-

sions: reliability, comparability, and sustainability, with specific scoring criteria as follows:

Data Quality Reliability Evaluation assesses data source authority and disclosure status. Data source authority is classified as national (released by the National Bureau of Statistics and relevant ministries), prefectural (released by local governments or statistical bureaus), or departmental (annual reports from local unit statistical departments or unavailable data), with corresponding scores of 10, 8.5, and 7.0 respectively. Data disclosure status is categorized as public (directly queryable), partially public (requiring contact with relevant institutions), or non-public (unavailable), with scores of 10, 8.5, and 7.0 .

Data Quality Comparability Evaluation examines data application scope and production methods. Application scope includes all regions (nationwide applicability), verified areas (specific provinces or cities), and unverified areas (unclear spatial scope). Production methods comprise direct acquisition (officially certified results available), indirect calculation (requiring self-collection of officially certified data for computation), and theoretical estimation (calculation using non-certified data or literature). Corresponding scores are assigned accordingly .

Data Quality Sustainability Evaluation considers temporal range and collection methods. Temporal range is classified as 1-5 years (recent continuous data), 6-10 years, or 11-15 years of available data. Collection methods include entirely from published data, partially from published data, and from unpublished data, with respective scores .

This study employs typical sampling to select 24 indicators from the evaluation indicator system for quantitative scoring, though the method applies to the complete system. Data sources include the National Bureau of Statistics official website (http://www.stats.gov.cn/), China City Statistical Yearbook (2000-2015), China Environmental Statistics Yearbook (1998-2015) defined as national-level sources, Shaoxing City Statistical Information Network data (1996-2015) and Shaoxing Statistical Yearbook defined as prefectural-level sources, and Shaoxing Science and Technology Information Institute statistical reports (2010-2015) defined as departmental-level sources.

---

3. Metadata Tracking Method

Metadata is information about data—structured information that describes, locates, and characterizes information resources, including data content, conditions, and other relevant features [10]. Metadata research originated in digital library management and search engine development to address explosive data growth [11-12]. As a recorder behind the data, metadata excels at diversely presenting data connotations. A primary reason for current low evaluation data quality is inadequate attention to effectively recording and managing critical

information behind the data. Metadata serves the fundamental function of helping to record data sources, transparentize collection processes, standardize data quality, and ensure data authenticity and accuracy. Metadata theory offers an effective solution for enhancing data quality reliability, comparability, and sustainability by tracking key metadata for urban sustainable development evaluation indicators.

---

4. Research Results and Analysis

4.1 Current Status of Assessment Data Quality

Data quality issues directly or indirectly affect assessment results, manifesting in three primary areas:

Insufficient Data Reliability: Urban sustainable development capability assessment data originates from numerous departments, with considerable non-public data requiring coordination with data-producing agencies. The absence of systematic recording of acquisition information hinders timely traceability and correction of data problems, impeding assessment progress and complicating future data collection.

Insufficient Data Comparability: China' s vast territory and diverse city types lead to varying statistical calibers for the same indicators across different climate zones and functional city categories. The lack of documented data production methods results in incomparable data being used in assessments, hindering comparative research across cities and comprehensive sustainable development strategy implementation.

Insufficient Data Sustainability: The long-term nature of sustainable development capability assessment requires continuous data acquisition. Temporal discontinuities caused by untraceable historical data severely impact assessment scientific validity and accuracy. Without persistent and complete data recording across diverse experimental zones and multi-source data producers, data authenticity and accuracy cannot be guaranteed, affecting result applicability. The absence of expanded data quality connotation in current acquisition and management mechanisms represents the root cause of these issues.

4.2 Comparative Analysis of Data Quality After Metadata Tracking

Applying the aforementioned scoring criteria to selected indicators before and after tracking key metadata items reveals significant improvements. Data quality scores for reliability, comparability, and sustainability all increased substantially after tracking corresponding metadata items . The overall score improvement stems from metadata' s enhancement of data source information, statistical methods, and caliber information.

Regarding **source information**, recording data-producing units (including institution names, contacts, phone numbers, addresses, and emails) facilitates timely traceability and error correction when issues arise, improving reliability. **Basic information** such as data lineage descriptions further supplements evaluation data with calculation formulas and data sources, enhancing comparability. **Statistical method and caliber information**, including statistical frequency, time, and spatial range, ensures comparability across different regions. Tracking key metadata items comprehensively improves data quality, demonstrating that systematic metadata recording and management provide indispensable advantages for long-term data quality assurance. Institutionalizing metadata tracking ensures stable, long-term recording, potentially exceeding sustainability score limits to meet the long-term requirements of urban sustainable development assessment .

Statistical analysis confirms these improvements. T-test results show that metadata-supported assessment data achieves significantly higher reliability scores (average increase of 0.87 points, $p<0.01$), with all indicators scoring above 9.0. For comparability, new scores average 0.55 points higher than original scores ($p<0.01$), with standardized statistical calibers unifying data production across Shaoxing' s departments. For availability, new scores show the most significant improvement, averaging 1.14 points higher, demonstrating metadata tracking' s predictability for long-term data acquisition. Overall, the comprehensive score increased from 8.50 to 9.33 (9.8% improvement), with reduced standard deviation indicating more stable quality.

---

5. Urban Sustainable Development Capability Assessment Metadata Management System

The advantages of improving evaluation data metadata are evident. This study developed a metadata management system based on the 14 key metadata items from the Science and Technology Support Program project. The system provides efficient and convenient metadata recording and management to ensure data quality in urban sustainable development capability assessment.

The metadata management system employs a three-tier B/S (Browser/Server) architecture: user interface layer, business logic layer, and data access layer. Developed following software engineering methodologies and processes [17-18], the system provides a user-oriented metadata interaction platform offering solutions for data providers, auditors, and senior managers [Figure 2: see original paper].

The system implements a user hierarchy with three roles: data provider, data auditor, and final approver, with progressively increasing management and retrieval permissions. The final approver holds highest-level authority, while data providers can submit and manage metadata for their affiliated cities. This hierarchy aligns with standardized government administrative processes, facilitates

accountability in data auditing and management, and ensures strict metadata quality control [19]. User classification also supports system development, as modular and hierarchical functions enhance data security protection and metadata database management [20].

Based on system functions and reporting requirements, the urban sustainable development capability assessment indicator system requires reporting against 14 key metadata items. Table 6 demonstrates sample reporting for the "Total Ecosystem Service Value" indicator .

---

6. Conclusions and Outlook

Urban sustainable development capability constitutes a complex system comprising economic, resource, and environmental elements. Establishing an evaluation indicator system that reflects these components and measures urban sustainability status is fundamental to assessment. Data quality is critical for scientific and effective results, as data fills the indicator framework. Thoroughly analyzing each indicator' s source, generation method, spatial and temporal scope, and urban development attributes is essential for accurately understanding sustainability.

Metadata theory and technology provide effective methods for addressing evaluation data quality issues. This study's results demonstrate that using a metadata management system facilitates metadata management and recording, thereby ensuring data quality and improving assessment result validity. China' s urban sustainable development capability assessment should comprehensively adopt metadata management technology, connecting it with sustainable development experimental zone systems to ensure traceability for every indicator. Future efforts should strengthen legal frameworks emphasizing metadata information, enhance awareness among government departments about metadata importance, and promote seamless integration between metadata management systems and sustainable development assessment platforms to produce more credible results.

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