Research on the Coupled Representation of Model and Data Integration Guided by Metadata Sets

Yucai Li [1], Liang Huo [1], Wangzhang Zhu [2], Peng Bao [3], Miao Zhang [1], Binghan Li [1]

[1] Beijing University of Civil Engineering and Architecture, Beijing, China - liyucai1211@163.com, huoliang@bucea.edu.cn, (zm000419, lbh292419)@163.com
[2] Geovis Technology Co.,Ltd, Beijing, China - zwzhlj@126.com
[3] CIGIS (CHINA) LIMITED, Beijing, China - baopeng0806@163.com

Keywords: Model and Data Integration, Metadata Description Expression Framework, Metadata Sets, Digital Pipelines

Abstract

Aiming at the problems of islands, redundancy and sharing difficulties of spatial and temporal business models and data in geospatial and spatial business systems, this paper proposes an integrated and coupled expression method of models and data guided by metadata sets, based on the research on core metadata of information resources and component-based encapsulation. First of all, this method classifies the model and data features, and constructs the metadata description and expression framework dominated by feature meta-components; then, establishes metadata sets through the expansion of the metadata description and expression framework; then, constructs the correlation between the model and the data through the use of the established metadata sets by means of the component-based encapsulation method, so as to complete the integrated coupling of the model and the data; and finally, realises the integrated coupling expression of the model and the data in the digital pipeline business system. The results show that this method integrates the expression of digital pipeline element models and data, solves the problems of silos, redundancy and conflicts of heterogeneous models and data from multiple sources, ensures the integrity of the attributes of the models and data elements, and promotes the structured expression and application of the models and data.

1. Introduction

Urban pipelines play a vital role in modern economies and societies. It constitutes the nerve and circulatory system of the city. However, with the acceleration of urbanisation and the expanding scope of social infrastructure, pipeline systems are facing more and more challenges, such as high system load factor, data silos, under-utilisation of resources, etc. It is a major challenge to carry out demand resolution in a fast and efficient manner.Geospatial models have become an irreplaceable and important tool in modern geographic analysis and research, as they provide a scientific basis for revealing the intrinsic mechanism of geographic phenomena and predicting the development trend of geographic systems. The operation of geospatial models requires the support of data, and the preparation of data for geospatial models that meet the requirements of application scenarios is the most basic work in the use of models. Due to the complexity of the research object itself and with the depth of the research, geospatial models are becoming more and more complex and require more and more data, and consequently, the model and the data are siloed, a large number of redundancy and occlusion and other problems (Li Pengfei, 2016), in order to improve the efficiency of the industry and the reuse rate of models and data, there are currently network data retrieval (Wang Yanjun, 2013), semantic reasoning (Miao Lizhi, 2015), automatic recommendation (Zhang Yuhang, 2022), and the establishment of a unified data representation and exchange model (Tao Hong, 2018) to solve the matching of models and data, but did not solve the problems of data redundancy and duplication and sharing (Jia Ping, 2021), several methods The main idea is to judge the matching based on the similarity between the two, but only based on the similarity can not reflect the specific differences between the data and the requirements of the type and location, it is difficult to carry out effective matching and combination, which directly affects the efficient matching of the model and data.

The IEEE First Metadata Workshop provided several definitions of metadata, noting that standardisation of metadata can better advance the process of data sharing and servitization. The metadata standards on geographic data that have been published so far are Content Standard for Digital Geospatial Metadata, CSDGM in the United States, GDDD Dataset Description Method in Europe, Geospatial Dataset Description in Canada, Core Metadata Elements in Australia, Geographic Information Metadata in China (GB/T19710-2005), and other metadata standards related to geographic data. Data (GB/T19710-2005) and other metadata standards related to geographic data, aiming to promote scientific and technological innovation, promote the standardisation of metadata, and provide a unified framework for the sharing and standardisation of geographic information resources (Pei Xintong, 2021).

The advantages of metadata for model and data coupling as seen through the nature of metadata and existing standards include improved data discoverability, understandability and interoperability, facilitating data quality management and reusability, and at the same time, helping to ensure data security, thus promoting the effective application of models and data in a variety of application scenarios (Ma Jong Won, 2021), while the establishment of metadata architecture frameworks in the corresponding domains meets the needs of integration and sharing of information, providing data resources and information systems for the effective application of information (Zhao Junxi, 2013). Therefore, this paper adopts the method of establishing model and data metadata description and expression framework to expand the construction of metadata sets, guiding the establishment of the association relationship between the two, and using this as the basis for constructing the model and data integration components and accessing the real-life three-dimensional system, thus realising the coupled expression of model and data integration.

https://doi.org/10.5194/isprs-annals-X-4-2024-217-2024

2. Metadata Set-Guided Model and Data Integration Component Building

The model and data integration expression model proposed in this paper is a componentised encapsulation model guided and constrained by metadata sets under the component-based idea. Firstly, the model and data of digital pipeline elements are classified by features, then the metadata description expression framework is established to organise and guide the construction of metadata sets, and then the established metadata sets are used to construct the association between the model and the data through the component encapsulation method, so as to realise the completion of the construction of the integrated model. Figure 1 below shows the overall technical flow chart of this paper.

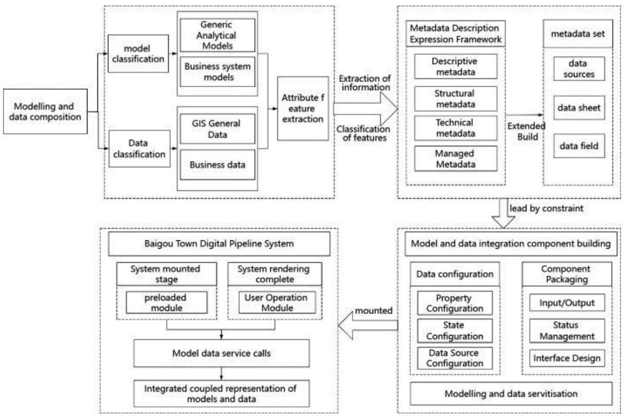


Figure 1. Technical flow chart

2.1 Model and Data Classification and Characterisation

Under the condition of GIS combined with specific domains, the models can be divided into two categories, generic analysis models and professional application models, according to their scope of application. General analysis model is the model geospatial model in the common analysis model is extracted separately to form a general analysis model class, the content of which is based on the GIS analysis model, including but not limited to siting analysis, buffer analysis, visual domain analysis and so on. Professional application model refers to the model applied to specific decision-making problems in specific fields, such as forecasting model, planning model, economic model, etc. These models are determined by specific business departments in accordance with the actual needs, and their contents also show the professionalism and applicability in specific fields.

The features of the digital pipeline element model are the user's summary of the model's functions in the pipeline business domain, including label features, state features, structural features, service features and accessibility features. Label features are used to describe the quick discovery of the model and also the functionality of the model. State features are used to describe the operating environment of the model and the performance of the model. Structural features are used to describe the algorithms involved in the model, the data requirements and parameter requirements for operation. Service features are used to describe the service information of the model, which describes the relevant information of different service types in a unified way, and the types include network services and local services. Accessibility features are used to describe the copyright agency of the model, specifying the conditions and platforms for use.

There are more ways to classify data, the geographic information industry combined with the field of urban pipeline systems can be divided according to the application of data can be expressed as general data and business data. Generic data are geographic information data that are universally applicable to a wide range of industries and domains. Business data is related to the business operations and needs of a specific industry, organisation or domain, and contains information in a specific business context, which is generated in the daily operations of an organisation and used to support business decisions and processes. In practice, generic data and business data are intertwined. For example, business data may involve generic time, location, and other information.

The features of digital pipeline element data are descriptions of the nature and behaviour of pipeline data sets, including label features, state features, structure features, content features and accessibility features. Label features are used to describe the identification information and category information of the data, which is also the visual representation of the data name. State features are used to describe the information source and spatial characteristics of the data, and express the distribution and changes of the data in the spatial dimension. Structural features are used to describe the internal organisation and arrangement of data in a particular pattern or form. Content characteristics are used to describe the specific information and attributes contained in the data, including the constraint rules of the data content, associations between data tables, quantities and value domains, and so on. Accessibility characteristics are used to describe the data's belonging information, management information and application information, which is the expression of the data's confidentiality, age, periodicity and other information.

2.2 Metadata Set Construction

The Metadata Description Expression Framework provides an efficient organisational way to manage and understand its key data and model information. The framework allows for the documentation of the model's algorithms and parameters, as well as key information such as the type and structure of the data. This metadata description representation framework helps to improve the comprehensibility and accessibility of the data, enabling business users to understand the purpose and effect of the model more easily.

The features of the model and data mentioned in the previous section, together as the underlying descriptive meta-constructs, form the middle layer of the reusable metadata description representation framework. Starting from the business requirements of GIS and live 3D business, the middle layer of the metadata description expression framework can be expanded according to specific features, which is specifically expressed as a complete multi-featured and multi-information metadata description expression framework as shown in Figure 2 below.

Figure 2. Schematic representation framework for metadata description

When establishing metadata sets, the core metadata of China's national information resources must be used as a reference standard, in line with the recommended standards for metadata set out therein. The national information resource core metadata standard specifies the core metadata items recommended for use in the process of publishing information resources, as well as the metadata attributes, defines the metadata, and puts forward the principles and methods for expanding the core metadata. Among them, the composition of core metadata of information resources contains 4 mandatory metadata elements - identifier, name, keyword and access restriction; 2 mandatory metadata entities - responsible person and information resource classification; and also includes 1 optional metadata entity -date; and 2 optional metadata elements-summary and information resource link address, for a total of 9 metadata items. By summarising the existing geospatial-related metadata standards, this paper establishes the metadata set of pipeline element models and data from the perspective of the integrity expression of both models and data, targeting the business needs of urban pipelines, according to the metadata description expression framework written above. The full set metadata set contains the aspects of belonging components, belonging tuples, belonging composite types, belonging basic types, Chinese name of the full set metadata items, English name, summary profile and data types. Tables 1 and 2 show some of the contents of the full-set metadata set.

Components	dependent tuple	Basic type of affiliation	Full set metadata item name	Overview	data type
Label Features	marking information	identifiers	modelName	Model name	Text
Label Features	marking information	identifiers	modelID	model unique identifier	Text
Label Features	marking information	classifier	modelType	Type to which the model belongs	Category
Label Features	signature information	model capability	function	Functions of the model	Text
Label Features	signature information	Areas of application	domain	Areas of application of the model	Text
Label Features	Spatial information	versatility	isGeneral	Is the range of model applications generic	Boolean
Label Features	Spatial information	application area	applicationRegion	Area of application of the model	Text

Table 1. Example of labelled features for a model metadata set

Components	dependent tuple	Basic type of affiliation	Full set metadata item name	Overview	data type
Structual Features	Parameter information	resolution	resolution	Resolution of data	complex
Structual Features	Parameter information	volume of information	size	Disk usage for data	Quantity
Structual Features	Format information	data format	format	Format of the data	Category
Content Features	Rules information	restrictive rule	rule	Rule constraints on data	Text
Content Features	Rules information	Summary of data	dataSummary	Summary definition of data	Text
Content Features	Related information	Data Association Keys	primaryForeignKey	Associated characters between data	complex

Table 2. Examples of structural features and content features of data metadata sets

2.3 Model and Data Integration Component Building

The metadata set defines the structure of the model and data as well as the rules, which provides the basic information framework for the construction of model and data integration components in this paper. This paper adopts the idea of component-based encapsulation to encapsulate the model in accordance with the established interface design, and access the system and express the model and data integration through the component-based encapsulation module components.

On this basis, the overall project component framework is firstly constructed according to the requirements of the business system and the contents expressed in the metadata set. For the core part of the component structure, the construction is defined vertically with reference to the structural features and service features of the model metadata set, including the construction of algorithms for individual models, method function definition, return value design, etc. Then the horizontal expansion of the core content of the structure is carried out according to the structural features, content features and accessibility features of the data metadata set, including data standardisation, algorithm data access reconstruction, parameter reconstruction, etc. The core part is shown in Figure 3. At the end of the definition of the core part of the component structure, it is introduced into the entry file, and the entry file is defined in the package configuration file, including the definition of the version number, license, keywords, dependencies, etc. At the same time, the state features in the model and data metadata set are used to supplement the test file, and the label features are used to supplement the package test file and the description file to complete the packaging of the component, and the access to the system to achieve the integrated application and expression of the model and data. The application and expression of model and data integration.

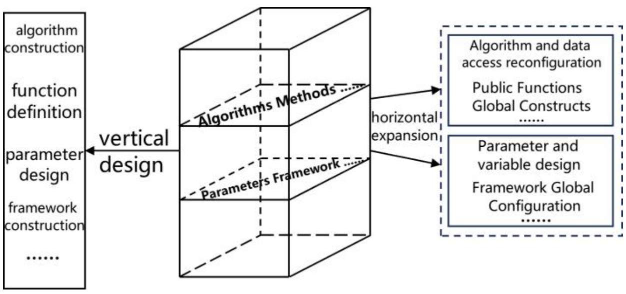


Figure 3. Schematic design of the core part of the component structure

3. Metadata Set-Guided Coupled Representation Approach for Model and Data Integration in the Baigou Digital Pipeline System

3.1 Models and Data Applied in the Baigou Digital Pipeline System

In order to verify the validity of the method of this paper, WebGL technology is used to draw the 3D scene of the BaiGou digital pipeline, and the method introduced above is applied to this real-life 3D system, and some of the models and data involved in the system are shown in Table 3 and Table 4.

Model type	Model Name
GIS Generic Analysis Model	CrossSectionalAnalysis Buffer analysis
Digital Pipeline Professional Application Model	ClosedValveAnalysis

Table 3. Model types

Data type	Data Name	Data Volume
GIS General Data	tilt model data	109GB
	Terrain data	1.25GB
	Remote sensing image data	27.1GB
	Point cloud data	37.6GB
Digital pipeline business data	Digital pipe network model data	74.5GB
	Digital Pipeline Entry Class Data	203MB

Table 4. Data types

3.2 Application of an Integrated Coupled Model and Data Representation Method Based on Metadata Set Guidance

In the digital pipeline system, the model and data are called by the system function, and the model and data integration coupling application is expressed in two aspects, which are the system preload function part and the user operation function part. The system preload function occurs in the system mounted stage, this stage instance system has completed all the processes such as data observation, compilation and rendering, creation of DOM, etc., and DOM operation can be carried out, and the model and data integration coupling expression application realisation of the part of the preload function occurs in this stage. The model and data integration coupling expression application of the user operation function part occurs after the system loading and rendering is completed, which is to capture the data space range and map scale that the system currently belongs to, and realise it according to the selected function. The overall application implementation flowchart is shown in Figure 4.

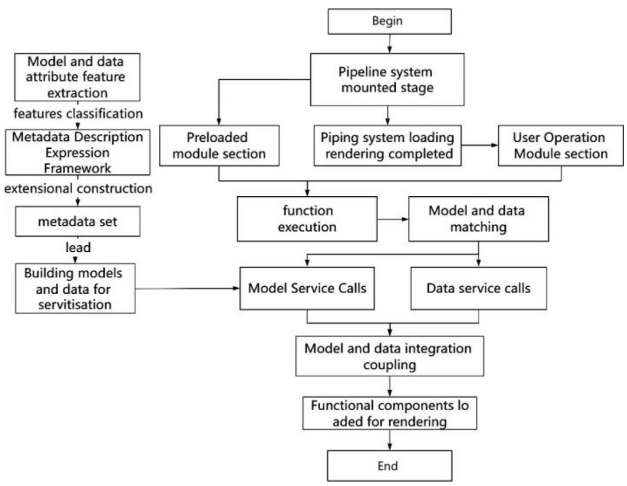
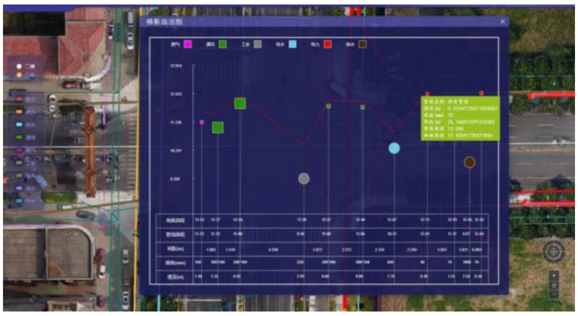


Figure 4. Schematic diagram of the application of the coupled model and data integration representation approach

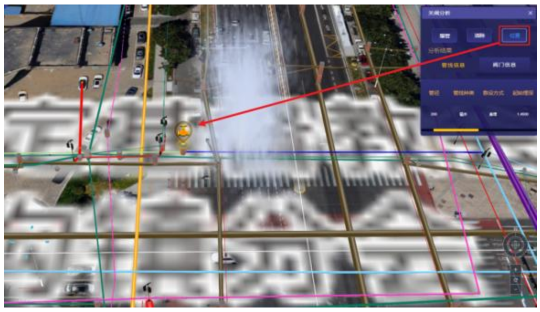
According to the above approach the matching models and data in the pipeline system are applied by means of the system functionality componentisation, and Figure 5 illustrates the application of some of the model and data integration representations.

(a) System Preload Module



(b) CrossSectionalAnalysis Module

(c) BufferAnalysis Module



(d) ClosedValveAnalysis Module

Figure 5. Visualisation of the integrated coupled representation of model and data

The response rate of each module of the system is obtained through the system console and this is used to verify the effectiveness of the method for model and data integration coupling. According to this statistical result (shown in Fig. 6), it can be seen that the research method of this paper makes the system's response to model and data integration tend to stabilise in a shorter time, and as the system running time is prolonged the impact on the response of model and data integration components is zero, which indicates that the application and expression of this method in the real-life three-dimensional business system has a strong applicability and is able to improve the user experience in a better way.

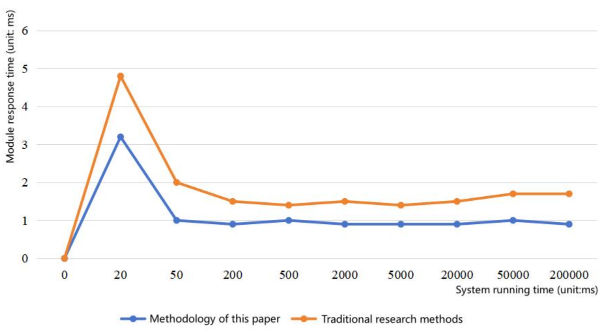


Figure 6. Line plotting of module response times

4. Concluding remarks

This paper mainly focuses on the application of GIS real-view 3D system, and based on the core metadata of information resources, component-based encapsulation and other related researches, it constructs metadata sets for business system models and data, establishes the correlation relationship between the two, and then realises the integrated coupling method of models and data by means of servitization and componentization. The information framework expressed in the metadata set regulates the information characteristics of the model and the data and solves the redundancy problem caused by cross-mixing of the two, and solves the silo and sharing difficulty of the model and the data by the servicing and componentisation method, and finally proves the scientificity and completeness of the method by the experiments.

The research results show that this method establishes a complete metadata set through feature extraction of GIS business system models and data, which leads to the integrated coupling expression of models and data. Meanwhile, it has application value in three kinds of GIS systems: terrain level, city level and component level.

Acknowledgements

This study was funded by the Shanxi Provincial Key Research and Development Programme: Research on key technology of spot extraction from satellite images of domestic satellites in Taiyuan urban area, Shanxi Province. (Project No. 202202010101005).

References

Diamantini Claudia, Giudice Lo Paolo, Potena Domenico, Storti Emanuele, Ursino Domenico., 2020: An Approach to Extracting Topic-guided Views from the Sources of a Data Lake. Information Systems Frontiers, 23(1), 1-20.

GB/T 41455-2022, Underground pipeline element data dictionary [S].

Guo Li, Jiang Jingli, Li Hao, Wang Yunge., 2020: Design of multi-source geospatial vector data association model.

Surveying and Mapping Bulletin. doi.org/10.13474/j.cnki.11-2246.2020.0222.

Jia Ping, Yang Jie, Zhu Yunqiang, Song Jia, Li Chenrong., 2022: Research on automatic matching method of geospatial model input data. Mapping and Spatial Geographic Information, 45(05), 26-31.

Jünger Stefan, Borschewski Kerrin, Zenk-Möltgen Wolfgang., 2019: Documenting Georeferenced Social Science Survey Data: Limits of Metadata Standards and Possible Solutions. Journal of Map & Geography Libraries, 15(1), 68-95.

Li Pengfei., 2016: Research on Heterogeneous Geospatial Model Sharing and Interoperability Methods, Wuhan University.

Li Rui, Lin Yanping, Xu Zhengquan, Feng Wei., 2011: Metadata scalability management for spatial data storage objects. Computer Application Research, 28(12), 4567-4571.

Ma Jong Won, Czerniawski Thomas, Leite Fernanda., 2021: An application of metadata-based image retrieval system for facility management. Advanced Engineering Informatics, 50: 101417.

Miao Lizhi, Xu Jie, Zhou Ya, Cheng Wenchao., 2015: Deductive reasoning for OGC geographic information services with application of description vocabulary approximation. Journal of Surveying and Mapping, 44(09), 1029-1035+1062.

Pei Xintong., 2021: Research on metadata standards and application of ontology-based geographic open data. Dalian Maritime University. doi.org/10.26989/d.cnki.gdlhu.2021.000732.

Pons X, Masó J., 2016: A comprehensive open package format for preservation and distribution of geospatial data and metadata. Computers and Geosciences, doi.org/10.1016/j.cageo.2016.09.001.

Shi Haibo., 2019: Design and research on surface coverage semantic metadata sharing system. China University of Mining and Technology.

Tang Lei, Hu Yidong, Meng Cheng, Zhou Haiyan, Liu Chengwei., 2023: Research on data classification system of urban information model under the goal of smart city. Surveying and Mapping Geographic Information, doi.org/10.14188/j.2095-6045.2021470, 48(05), 149-154.

Tao Hong, Le Songshan, Jiao Dawei, Sun Lingzhi, Lu Yucheng, Wang Lingshan, Zhen Peipei, Lv Guonian., 2018: A data organisation method for urban cell scenarios oriented to geoanalytical simulation. Journal of Surveying and Mapping, 47(08), 1080-1088.

Wang Yanjun, Xu Dijun, Li Danong, Shao Zhengfeng., 2013: Service-oriented architecture for online updating of geographic elements. Surveying and Mapping Bulletin, (06), 41-44.

Ye Aidong, Chen Jinlin, Zhang He., 2020: A model-driven spatial database quality checking platform construction. Surveying and Mapping Bulletin. doi.org/10.13474/j.cnki.11-2246.2020.0090.

Zhang Xin, Chen Yong, Hu Bengang, Zhangfang., 2022: Fusion of attributes of urban underground pipeline elements. Mapping and Spatial Geographic Information, 45(04), 199-203+208.

Zhang Yuhang, Zhou Xiaoguang, Hou Dongyang., 2022: A comprehensive multi-factor crowdsourcing task recommendation method for geospatial data. Surveying and Mapping Bulletin. doi.org/10.13474/j.cnki.11-2246.2022.0015.

Zhao Junxi, Hua Yixin, Zhang Yajun., 2013: Design of geospatial metadata system framework in digital boundary. Surveying and Mapping Bulletin, (10), 47-49.

Zhu Jie, Zhang Hongjun, Liao Xianglin, Tian Jiangpeng., 2021: Implicit semantic understanding of geo-environmental spatio-temporal data based on topic model. Journal of Surveying and Mapping, 50(10), 1404-1415.