# Building an Integrated Toponymic Information System Based on TGIS and Big Data Technology

Bai, J. T.[1]  Pan, W.[2*]  Hou, Y. J.[3]  Zuo, Y. Q.[1]  Yang, H. H.[1]

1. Xi'an Map Publishing House, Xi'an 710054, China;
2. Collage of History & Archives, Yunnan University, Kunming 650091, China;
3. Center for Historical Environment and Socio-Economic Development in Northwest China of Shaanxi Normal University, Xi'an 710062, China

**Abstract:** Storing toponyms and their historical changes by employing a scientific approach allows querying them for different historical stages and realizing their visual representation. Such a toponymic information system can be exploited for historical and geographical research by integrating temporal and spatial information utilized for demonstration and analysis tasks. Thus, this paper builds a toponym database by combining a spatio-temporal framework and big data technology with a Temporal Geographic Information System (TGIS). The proposed method supports updating and maintaining toponymic information and attempts to create a modern information technology that maintains and broadcasts toponymic information. Specifically, we built a toponym database including spatial and temporal information and an integrated toponymic information system that incorporates tools to analyze the information included, along with a toponym updating mechanism based on big data analysis. Additionally, our framework provides solutions for various research projects to reference or confirm each other's results, improving the extent and depth of public services based on the relevant results and promoting the integrated utilization of research results. Finally, our toponymic system affords revealing historical toponymic information and meeting the asymmetrical and complementary demands.

**Keywords:** TGIS; toponym; spatio-temporal framework; visualization; analysis

## 1 Introduction

Toponyms are classified as essential geographical information and are often used to access a more comprehensive information range. With the advent of the big data analysis era, toponyms have increasingly been accepted as a critical tool to access data resources for academic research, information query, and information interactions[3,4]. The rapidly evolving progress in several aspects of society, combined with the advancement of big data analysis, artificial intelligence technologies, and the easiness to access the internet, imposes quick updates of toponyms and massively transmitting toponymic data through the internet. Despite the toponyms' importance, their changes over space and time have not been

explored yet[1]. Therefore, toponym management and its related services have become challenging recently, as staffs responsible for collecting, updating, utilizing, maintaining, and managing toponymic data have to deal with several emerging problems[2].

Toponyms possess three essential features: space, attributes, and time associated with general geographical phenomena[4]. Temporal toponymic information is irregularly stored or lost due to incomplete records and historical data storage[5]. Nevertheless, mastering historical data of geographical objects allows analyzing current data and retrace the geographical objects' history to ultimately determine regularities and predict future development[6–9]. Spurred by the advantages of maintaining accurate toponymic information, this research introduces big data analysis technology to build an integrated toponymic information system. Specifically, we build a spatio-temporal toponym database by adopting semantic analysis and automatic update technology that relies on big data analysis and spatio-temporal data modeling theory. Furthermore, we set up an integrated toponymic information system acting as a tool for spatio-temporal analysis and as a mechanism to update toponyms.

## 2 Logical Framework of the System

A spatio-temporal toponym database based on big data technology and a Temporal Geographic Information System (TGIS) can provide a framework to establish an integrated information system[10,11]. The database shall provide researchers fundamental scientific data and the appropriate tools to search for toponyms and their relevant fields to deliver better social services. The system framework diagram is illustrated in Figure 1.

## 3 Structure of Spatio-temporal Toponymic Data

Considering toponyms being a type of geographical expression involving nature and human society attributes, automatic toponym updating has to rely on a spatio-temporal database of toponyms and an analysis of the spatio-temporal domains. Relevant data should be obtained from different fields, but different data demands across various industries and fields impose differences in historical information mining and description. These differences should be excluded, narrowed, or employed to confirm each other through semantic analysis and then act as determinants or supporting evidence during the maintenance and update of the historical toponymic information. The toponymic spatio-temporal domains often vary randomly, and thus, the spatio-temporal state and grid model have been proven quite effective in organizing and managing toponymic Spatio-temporal data[7]. To facilitate the spatio-temporal analysis and visual presentation of historical toponyms, we built a spatio-temporal database based on the GIS spatial topological relation theory, whose conceptual design is presented in Figure 2.

## 4 Data Processing

Generating the second-level district grids is the primary modeling procedure, which exploits modern information technology to store all toponymic spatio-temporal information and modern cartographic techniques to visualize the toponymic data. This modeling algorithm improves the toponymic information storage efficiency and affords information query and integrated analysis. We adopt the traditional administrative division to fully exploit the
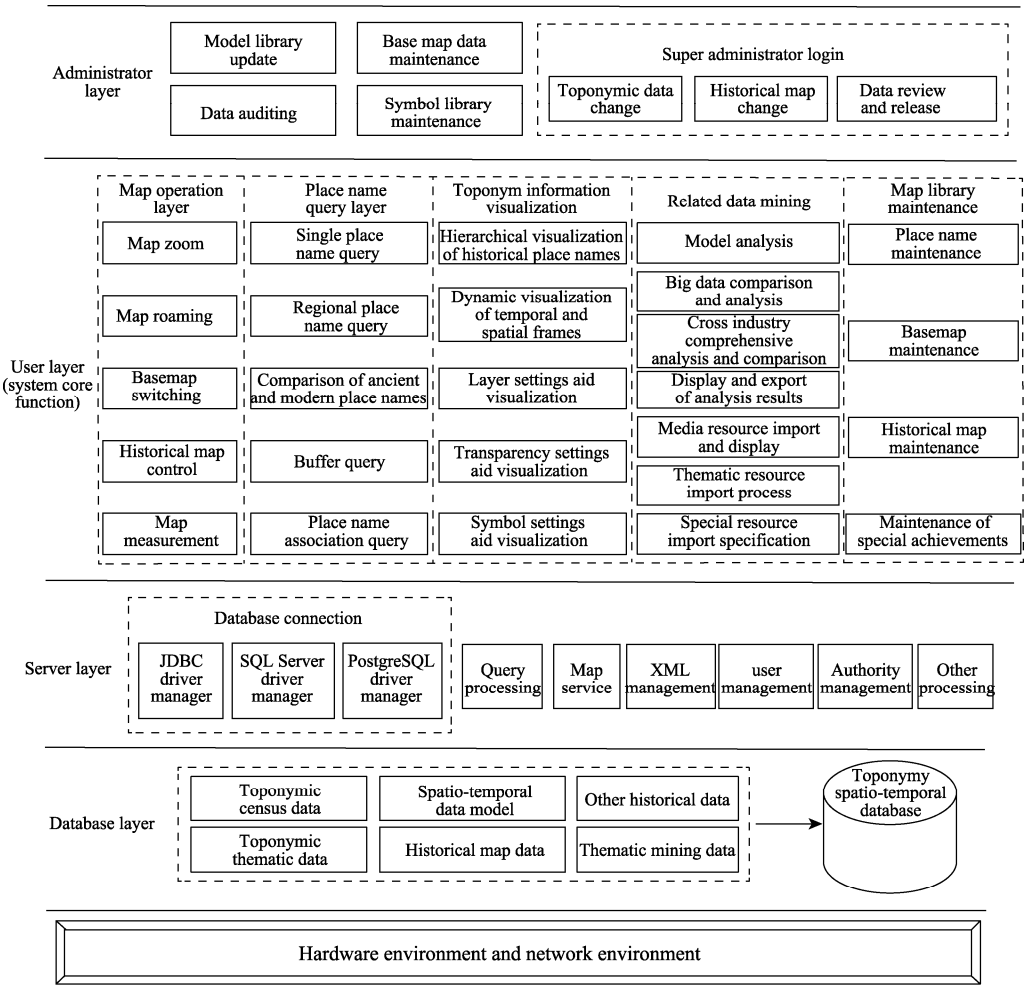
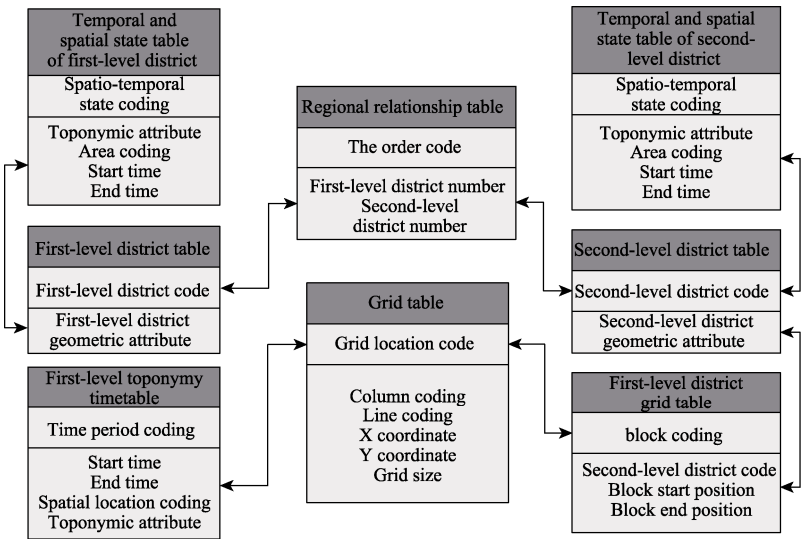**Figure 1**    Framework of the spatio-temporal toponym database



**Figure 2**    Flow chart of the database development

algorithm's advantages and build the basic expression units. Grids that are interconnected within an administrative district constitute a basic expression unit, for which the system only needs to record the position codes of two grids: the first and last ones. Grids are coded from left to right by row according to the attributes of toponyms and grid locations. Grids that are interconnected and have the same toponymic attributes constitute a district. If a grid is located at the boundary of the coding district, it will be classified into the district with the district's boundary being redefined. Otherwise, a new district is defined, and the grid will be used to delimit the boundary of the new district.

The core of the second-level district indexing is processing toponymic units represented by polygon vector data. The data processing results are second-level districts, which are then coded and recorded, and the logical connections between first-level and second-level grids are identified. Second-level districts are generated employing the Union tool to deal with administrative district units and district map layers. Every second-level district is then coded through loop computation, the logical connections between the first-level and second-level districts are identified, and their codes are recorded as the digital basis of the district relationship table. It should be noted that during this process, the attribute information needs to be maintained to calculate, identify, label, store, and rank the second-level districts' grid coordinates.

## 5 Analysis, Statistics, and Processing of Spatio-temporal Information of Toponyms

The spatio-temporal analysis mainly includes locating toponyms in space, along with toponymic extraction, merging their temporal sequences, calculating their frequency statistics at specific locations, calculating their accumulative lifetime, and generating their spatio-temporal volume matrix.

### 5.1 Locating Toponyms in Space

When a new toponym is captured by big data analysis technology, its spatial location is identified based on its spatial attributes. Then the toponym's temporal attributes and stories are employed to confirm whether it is a new toponym. If it is a new one, a new toponym record should is set. Otherwise, it is defined as historical information of an existing toponym and thus is stored under the previous toponym.

### 5.2 Toponym Extraction

According to relevant standards and regulations, toponyms of administrative districts generally comprise two names: proprietary and standard. Most databases store complete toponyms of administrative districts, while the proposed algorithm only extracts proprietary names and records them as toponyms.

### 5.3 Merging Time Intervals of Toponyms

Changes of toponyms related to administrative districts commonly involve their common names, e.g., "township" changed into "town", but proprietary names are rarely altered. For toponyms sharing the same proprietary name, if they refer to the same place, their time spans should be merged, and the toponym should be stored in the toponymic record. The process of information merging is as follows: temporal attributes and common names of coded grids are recorded as the time field and are then ranked. Adjacent grids sharing the same toponym whose time is successive have their records merged, with the start and end

times of the toponyms determined based on the time sequences of the grids. If time is ambiguous, then a synchronous record is created, while if the adjacent grids have no such connections, the next record will be checked.

## 5.4 Statistics of Toponym Frequency of Specific Locations

A specific grid may include several historical toponyms, potentially being repeatedly adopted from time to time. All grids are sequentially dealt with, and all grid records are stored in a timetable with the toponyms ranked based on their starting time. If one toponym has not yet been included, it is added to the storage table. Otherwise, it will be dealt with during the following time interval. This process continues until all grids are considered. Meanwhile, the boundaries and the total number of toponyms per grid are recorded to form a matrix comprising grid locations and the total numbers of toponyms. The toponyms' changes related to target districts and their coverage are revealed by combining the initial grid data.

## 5.5 Toponym Lifetime

According to the toponyms' spatio-temporal information statistics, the number of toponym repetitions, the lifetime of each repetition, and the cumulative lifetime of one toponym are three significant indicators revealing the historical, cultural, and regional changes of a place. The lifetime of a toponym refers to the time duration of a toponym's spatio-temporal state, with the number of repetitions and accumulative toponym lifetime involving more complex definitions. The accumulative lifetime refers to the sum of the same toponym's repetition time intervals at the same location. If toponyms are interrupted, the start and end time of a toponym cycle should be recorded and the number of interruptions determines the number of toponym repetitions. Nested loop tools can be used to deal with every toponym and its spatio-temporal state.

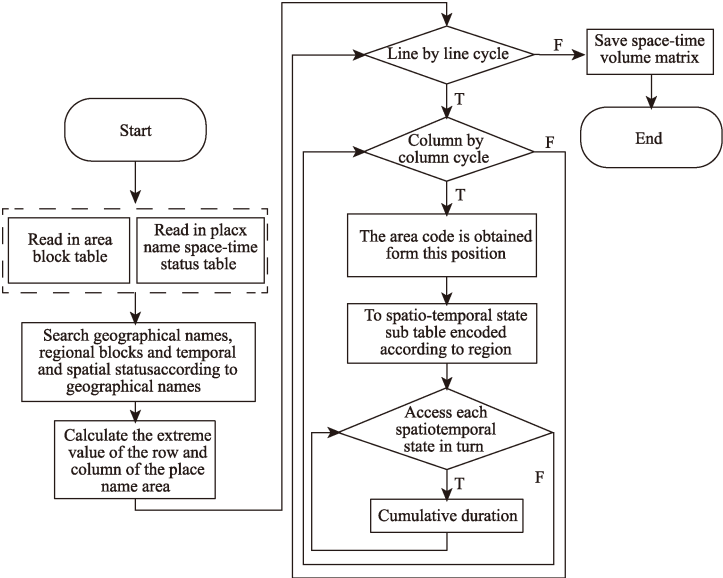## 5.6 Generation of Spatio-temporal Volume Matrix of Toponyms

The spatio-temporal volume matrix involving the spatial coverage of a toponym is the primary target of the toponyms' spatio-temporal analysis. The spatio-temporal volume matrix pipeline is illustrated in Figure 3. According to the district classification table and the toponym spatio-temporal state table obtained through toponym retrieval, the spatio-temporal state sequence and district coverage can be obtained. Then the row and column limits of the toponym's spatial coverage are calculated through nested loop processing, and the spatio-temporal state of every grid is retrieved and added to the accumulative lifetime of the toponym. The spatio-temporal volume matrix of the toponym is obtained when the cycle is completed.

# 6 Visualization of Changes of Spatio-temporal Toponymic Information

Given the known toponymical data characteristics, semantic analysis exploits big data analysis information acquired from different industries. Hence, combining the semantic analysis results with a toponym's historical information can confirm whether the toponym is newly added, altered, removed, or disappears and affords to obtain the spatio-temporal change process and toponym's state. To visually represent the dynamic process statically, several designs are available to represent spatial states, including dot, line, and plane-type toponyms based on a selection of visual variables such as shape, size, and color.

Considering that a toponym change often lacks regularity, static visualization can better describe the spatio-temporal changes of toponyms, with the latter classified into three
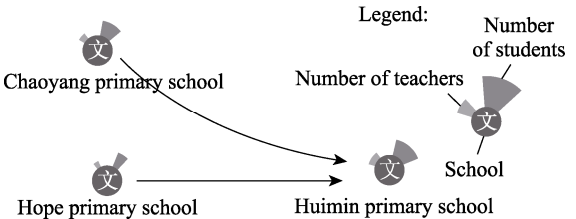
categories according to their spatial states: dot, line, and plane. To represent toponyms, visual symbols should be selected to present shape, color, and size variables. The similarities and differences before and after the toponym changes should also be considered during the design to highlight this dynamic process.



**Figure 3**    Flow chart of generation of spatio-temporal volume matrix of place names

## 6.1 Visualization of Spatio-temporal Changes of Dot-type Toponyms

Toponyms may change concerning two aspects: spatial location and attributes. Spatially, toponyms involve three types of changes: addition, cancellation, and transfer. The first two change types are easily denoted through symbols, while for the third one, lines are used to indicate the locations before and after the "transfer" and are combined with different colors to create a visual hierarchy. These are easier visualized regarding the attribute changes, as the toponym grade change is generally represented by altering the symbol's size or style. However, if changes are more complex, symbols are extended to enhance visualization and exhibit the toponym attributes' changes based on symbol alteration. An example is illustrated in Figure 4.
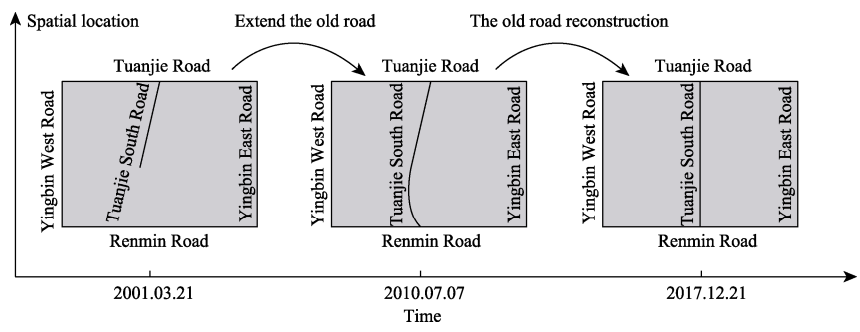


**Figure 4**    Example of spatio-temporal changes of dot-type toponyms

## 6.2 Visualization of Spatio-temporal Changes of Line-type Toponyms

In this work, we utilize lines or colors to visualize spatial location changes of the line-type toponyms. Figure 5 presents road location changes, mainly including extension and rerouting.

The line-type toponyms' attributes changes are more flexibly visualized utilizing different lines, colors, and line width combined with annotations (Figure 6).

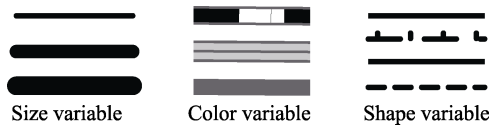**Figure 5**    Example of visualization of spatial changes of a toponym

## 6.3 Visualization of Spatio-temporal Changes of Plane-type Toponyms

Plane-type toponyms present changes in coverage shape, area, and attributes, represented by visual tools such as color, shape, and map layers. This paper focuses on the changes in administrative districts.



**Figure 6**    Example of visualization of changes of toponym changes

It is assumed that streets should fill one county-level administrative district, roads, towns, and residential communities without voids or overlaps in space, and administrative districts at the same level should be near to, border on, or separate from each other space.
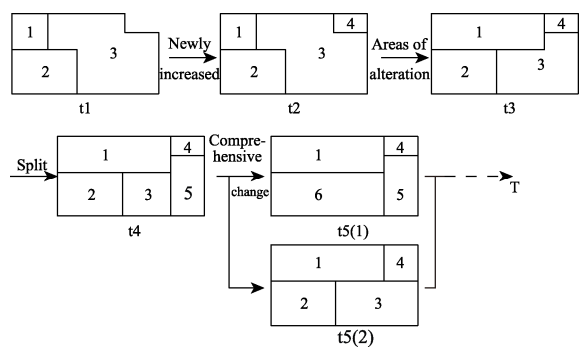
The spatial coverage of a plane-type toponym often changes in multiple directions, which are generally accompanied by the change of shape or hierarchy and its attributes. Visualizing plane-type toponyms is more complicated than line toponyms, and the effect of visualization varies considerably, which is closely related to the scale and size of the districts. Different colors represent administrative districts with different toponyms, and the hierarchy of toponyms is related to the temporal sequences of their adoptions: the later a toponym is adopted, the higher its hierarchy.



**Figure 7**    Example of visualization of location changes of plane-type toponyms

This strategy affords directly showing the state of a plane-type toponym in a specific time and its changing process (Figure 7).

Annotations or symbols can be added or altered in the mapping space to visualize plane-type toponym attribute changes. For example, as shown in Figure 8, gross domestic product (GDP) changes of agriculture, industry, and service sectors in a specific administrative district can be visualized.

When many attributes of a plane-type toponym change, such as shape and boundary, several visualizations should be combined to represent the complicated change, such as simultaneously adopting color and hierarchy. Hence, we adopt layer opacity to display the toponyms hierarchy as an auxiliary visualization method in this work. When maps of different time points overlap, and according to the user's discretion, some map layers can be displayed or hidden to demonstrate a hierarchy, can be shown by reducing or eliminating interruptions between map layers and expressing target information accurately. Through adjusting opacity, problems related to information expression due to multiple overlapping

layers are solved. More complicated changes can also be demonstrated through information charts. Figure 9 illustrates an example of a plane-type toponym visualization.

# 7 Functions of Integrated Spatio-temporal Toponymic Information System and Its Management

## 7.1 Major Functions of the Integrated Spatio-temporal Toponymic Information System

The database comprises nationwide or even worldwide fundamental and historical toponym data of Shaanxi province. Current historical data originate from national historical maps in the Historical Atlas of China and historical maps of Shaanxi province of different dynasties since



**Figure 8**   Example of visualization of attributes of a plane-type toponym

the Qin dynasty. The historical information employed includes 9000 pairs of Shaanxi ancient and current toponyms for comparison, 65,000 important heritage sites in Shaanxi province, more than 1,000 relics, nearly 200 traditional architectures, more than 100 ancient tombs, more than 30 religious and cultural sites, and nearly 10,000 locations mentioned in poems. All toponymic information is dynamically updated by experts who register the platform, with the update speed being subjected to the toponym management approval authorities and research achievement publication. The system is updated with the latest research results utilizing data from different research institutions.

This research adopts big data analysis technology to obtain research results of ancient toponyms in different industries and fields, ultimately matching ancient against modern ones. Then the matching results are comprehensively analyzed to extract the toponyms' changes and the stories related to each toponym and store them in a related database. The proposed architecture is illustrated in Figure 10.

The Spatio-temporal toponymical information database primarily comprises historical map data of
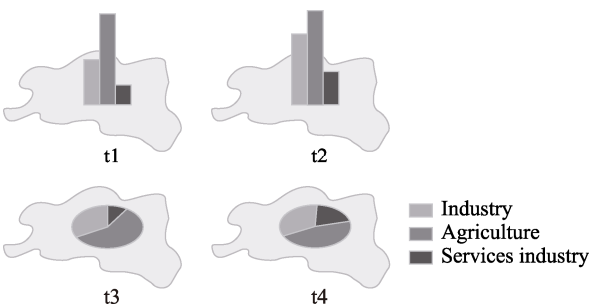


**Figure 9**   Example of visualization of overall changes of a plane-type toponym

well-known cultural cities in Chinese history, ancient map data, historical images, and spatio-temporal toponymic data of China. The database is based on technical regulations of data input and analysis and integrates a cloud model scheme. The database provides tools for online mapping and visual representation. The proposed architecture can be generalized to a platform demonstrating the results of a relevant field, exchange ideas, and manage data, and also it can be exploited by history and geography researchers for sustainable data mining and production based on GIS. The integrated spatio-temporal toponymic information system based on the database has the functions of collection, management, updating, analysis, visual presentation, downloading, and sharing of toponyms, meeting the requirements of various fields on query and utilization tasks related to reliable historical toponymic information. Some functions are presented in Figure 11.
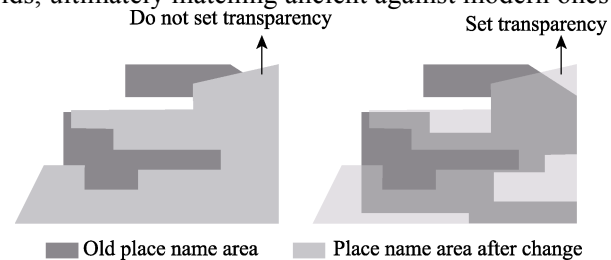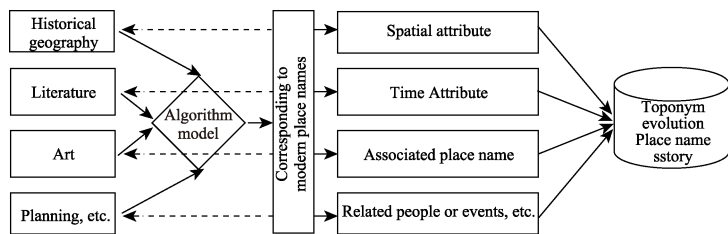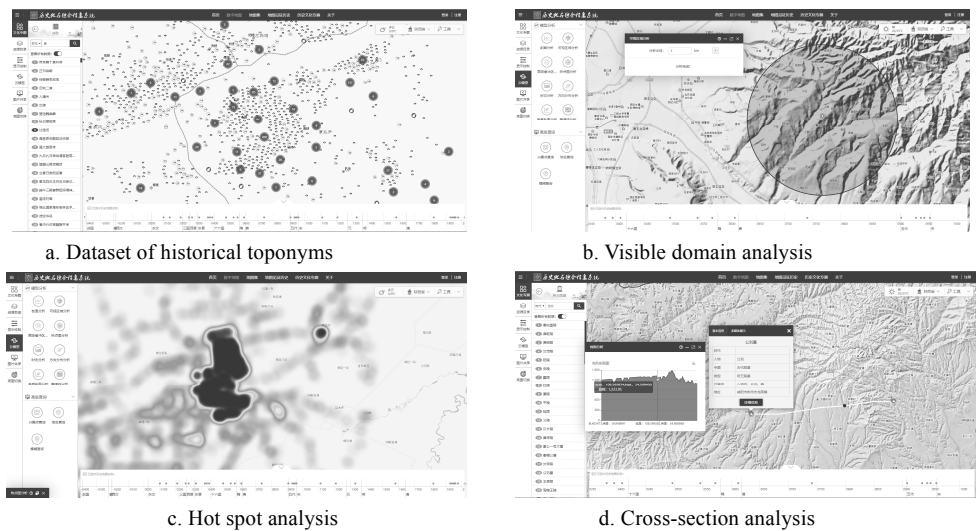
**Figure 10**  Application of big data technology in updating toponymical information



| a. Dataset of historical toponyms | b. Visible domain analysis |
|---|---|



| c. Hot spot analysis | d. Cross-section analysis |
|---|---|

**Figure 11**  Example display of functions of the integrated spatio-temporal toponymical information system

## 7.2 Management of the Integrated Spatio-temporal Toponymic Information System

This work is part of a project funded by the Ministry of Finance under the "2016 state-owned capital management budget for state-owned cultural enterprises", categorized as the state-owned economy structural adjustment expenditures with number 2230201. Xi'an Cartographic Publishing House is responsible for launching and maintaining the project, and participating organizations include the History and Geography Institute of Shaanxi Normal University, Xi'an International Studies University, Baoji University of Arts and Sciences, Capital Normal University, China Institute of Urban and Rural Construction and Cultural Heritage of Xi'an University of Architecture and Technology, Xi'an Civil Affairs Bureau, Xi'an Toponym Association, and Shaanxi Provincial Library. More organizations may be invited to participate in the project according to work demands.

## 8 Discussion and Conclusion

The proposed system allows updating and searching historical toponyms based on big data analysis technology and TGIS theory, revealing the value of toponyms and the relevant resources. Our method can easily be extended as a standard data and search tool for relevant researchers. Additionally, the results obtained can be edited into popular books to broadcast the related cultures and knowledge to the public online and offline.

The integrated spatio-temporal toponymic information system based on TGIS and big data analysis technology integrates toponym data acquisition, intelligent database

maintenance, static visualization of dynamic data process, toponym query and assistant analysis, and display and visualization of toponym data or the related results. The proposed intelligent toponym update theory based on big data analysis technology supports querying and displaying historical toponyms and is the core process for the spatio-temporal toponymic information framework and database presented in this work.

　　Future work shall involve expanding the database's spatial coverage to store more historical toponyms and along with their relevant data, improving the functions of intelligent matching ancient with current toponyms, developing toponym information indexing based on research results of ancient maps and literature, and creating a server to publicly broadcasting stories of toponyms and focusing on toponyms with cultural significance. Finally, data acquisition from toponym research within institutes or management departments should be gradually improved.

### *Author Contributions*
Bai, J. T. and Pan, W. designed the algorithms of historical place name comprehensive information system; Zuo, Y. Q., Yang, H. H. and others contributed to the data processing and analysis; Bai, J. T. designed the model and algorithm; Zuo, Y. Q. made data verification; Bai, J. T. wrote the paper.

### *Conflicts of Interest*
The authors declare no conflicts of interest.

## References

[1]　Huang, D. N. Building the query and management system about toponym temporal data based on TGIS [D]. Fuzhou: Fuzhou University, 2014.

[2]　Wang, D. P. Design and implementation of Guiyang digital place name public service platform [D]. Chengdu: University of Electronic Science and technology, 2010.

[3]　Ren, D. F., Xu A. G., Zhu, Y. J. Establishment of Fuxin place name query system [J]. *Geospatial Information*, 2011, 9(6): 107–110, 1.

[4]　Liu, L., Lu, J. S. Design and implementation of geographic name information touch query system based on GIS [J]. *Surveying and Mapping and Spatial Geographic Information*, 2012, 35 (10): 110–112.

[5]　Zhang, B. G., Wang, R. S., Gao, L. Spatio-temporal data model of point place names [J]. *Land and Resources Remote Sensing*, 2005(4): 82–85.

[6]　Zhou, Y. X., Fu, Z., Liu, D. W., *et al*. Study on temporal and spatial process of soil desertification, salinization and grassland degradation in Western Jilin province [J]. *Journal of Jilin University* (*Geoscience Edition*), 2003 (3): 348–354.

[7]　Fu, X. Q. Research on spatiotemporal state data model of historical geographical names [D]. Shijiazhuang: Hebei Normal University, 2019.

[8]　Smith, B. Engaging geography at every street corner: using place-names as critical heuristic in social studies [J]. *The Social Studie*s, 2018, 109(2): 112–124.

[9]　Choi, S. H., Wong, C. U. I. Toponymy, place name conversion and wayfinding: South Korean independent tourists in Macau [J]. *Tourism Management Perspectives*, 2018, 25: 13–22.

[10]　Zhao, S. H. Design and implementation of geographic name information integration based on 3S technology [D]. Wuhan: Wuhan University, 2017.

[11]　Yao, X., Zhu, D., Ye, S., *et al*. A field survey system for land consolidation based on 3S and speech recognition technology [J]. *Computers and Electronics in Agriculture*, 2016, 127: 659–668.

[12]　Ai. J. H. Research on place name address matching algorithm [D]. Kunming: Kunming University of technology, 2019.

[13]　Tan, Q. X. Atlas of Chinese History [M]. Beijing: China Map Publishing House, 1982.