THE ANALYSIS ABOUT SOLVING THE PROBLEMS OF DESIGNING AND COLLECTING METADATA OF THE HISTORIC MAP REFERENCE

Hongyun Song, Jiangqi Zhang, Ming Li, Jia Li

Affiliations:
Hongyun Song, female, senior engineer, master's degree, work in National Geomatic Center of China, mainly engaged in the collection and application of archival data of surveying and mapping geographic information.
E-mail：
songhy@ngcc.cn
Tel: 86-10-68417690
Fax: 86-10-68424073

Abstract
It is not an easy job to choose the historic maps from tremendous amount of maps all over the world, for a special project, such as "China's East China Sea, the South China Sea and neighboring countries historical map data compiled". The works include of, who are the manager of these maps, the managers record what kind of information about these maps, which of them can be used, and which of them need to supplement perfect. The location information of the historic map is very useful. According to demand of the project, the rules of designing historic map metadata and the reasonable historic map metadata item can be determined, the metadata information system is being built up, and all of those works improved the effectiveness of the historic map database.

Key words: historic map; archive references; metadata

INTRODUCTIONS

Historical Map is a kind of map that reflects the natural and political, economic, military, cultural conditions and their changes in human history. It is a map showing the spatial distribution and regional differences related to human activities in history. Historical Map Archive References, which generally refer to the maps that reflect the socio-economic geography of various periods in the history. A map on a critical time node can be a historical proof to witness a significant event. Each geographical name and each boundary line on the map would reflect the attribution of the region and the point of view represented by the cartographer in that time.

"China's East China Sea, the South China Sea and neighboring countries historical map data compiled" is a work project for the basic needs of national science and technology. The main task of the project is through collecting and organizing to analyze a large number of historical map data of the East China Sea, the South China Sea and the surrounding countries, modern map data and historical map data of different countries, which are kept in the national surveying and mapping archival data system and related departments. It provides the reference material and the surveying and mapping technology guarantee that have reference values for scientific research of the future political diplomacy and etc.

Four topics are designed in the project. Among them, the first topic collection and arrangement of maps about the area of the China's East China Sea, the South China Sea and neighboring countries and the fourth topic construction of historical map database in East China Sea, South China Sea and surrounding areas all involve metadata analysis, utilization and redesign of map archives. The first topic is to use the metadata of historical map data to collect and organize all kinds of map materials that are in line with the requirements. The fourth topic is to make metadata rules according to the new requirements with all kinds of map data collected and to carry on the input and management of the related information.

So how can we make sure the collection is useful and manage it the most scientifically, use it the most effectively? Starting from analyzing the metadata of all kinds of map files, we can find the metadata items of the rapid search of related map data. On the basis of analyzing the metadata and related standards of various existing map files, a new

metadata design for this project is carried out. Or to carry out the design of metadata that conforms to thematic requirements, establish metadata for thematic historical map data, and lay the foundation for the subsequent data building management and service delivery. All these are the problems to be solved in this paper.

1. Analysis and processing methods of metadata for different sources of map files

The source of the historical map data in this project is mainly divided into:

1.1 Analysis of existing material

It mainly includes the map resources accumulated by the national map archive data collection and management agencies, such as National Surveying and Mapping Archives, China Cartographic Publishing House and Naval Publishing House. It mainly includes: various scale topographic maps, charts, and teaching maps, used for the compilation of the map works of the classic various domestic and international atlases. The advantage of these map resources is that the map is a series, the compilation unit of map data is normal, and the quality of map is guaranteed. More importantly, different units have a certain management mode for map files. In order to facilitate the management and provision of services, some relevant metadata information about map data has been recorded. The definition and sufficient analysis of these metadata improve the efficiency of selecting the map data that meets the requirements, and some metadata information can be directly applied to some selected information. The following is an example of the management metadata analysis of the map data collected by China Map Publishing House and National Surveying and Mapping Archives.

Based on the content of a certain project, the relevant staffs of the map publishing house have checked the map resource one by one. Its metadata information is more detailed. The description of the information is mainly divided into 3 categories, 24 items of data description.

The publication information, including: name of publication, country of publication, publisher, date of publication, reprint edition, collector, book number, back cover scanning, copyright information scanning;

Basic information of the map, including: name of map, type of map, kind of text, original size (width×height, unit: cm), scale, date of mapping, compilation, type of data, map scanning The relevant information of the South China Sea, including: the scope of the region, the condition of the discontinuous line, the condition of the name of the island and reef, the condition of the ownership of the island and reefs, the favorable conditions and the description of the notes.

From the above information, it can be seen that the metadata information of map catalogued by China Map Publishing House focuses more on the record for the description information of map data, which is more convenient for material usage.

A large number of geographic information archives are managed by the National Surveying and Mapping Archives. The data collection patterns of different types of archival data are different. The metadata of the topographic map is divided into two modes in China and abroad. Among them, there are 39 metadata items for the publication of topographic data in China, including 12 of the archives management, accounting for 30% of the metadata. There are 27 items related to map data map attributes, accounting for 70% of the metadata items. Among them, there are only 4 items related to select map data, that is, the left longitude, the right longitude, the upper latitudes and the lower latitudes of the topographic map archives, which account for only 10% of the metadata items.

There are 37 contents of the metadata of the topographic data published abroad, and the use rate of the metadata in the rapid search is similar to that of the above.

It can be seen that the metadata information of National Surveying and Mapping Archives is biased to the content of archival data management, and focuses more on the input of management content. The purpose of the project is to quickly find the map data information related to the project in the collection of more than 300,000 metadata records. The quickest way is to make use of the record content of location information in metadata, and determine the scope of the map data, so as to determine the scope of the required information.

Information on the left longitude, the right longitude, the upper latitudes, and the lower latitudes of the metadata items recorded according to the topographic file data, the project group used the FME (Feature Manipulate Engine) software that is currently leading in the area of spatial data transformation, and realized the conversion of the metadata items recorded in each map to the spatial geographic information data. The spatial orientation of coordinates in ArcMap greatly improves the speed and accuracy of selecting the suitable map archives. By using the spatial geographic information data of the topographic map files, and with the help of geographic information software platform, it is more

convenient for the selection of the domestic and foreign published maps of map archives in accordance with the regional conditions in the related areas related to the subject study, drawing the graph of the distribution area of the data. The selection is according to the scope of the specific coverage of the data and the contents of the metadata of the other related topographic maps.

As for the data of the collection of special topics and atlases, a detailed interpretation of the relevant archives can also be selected through the area of the metadata item. Taking the overseas atlas for example, there are 29 contents of the existing metadata. According to the requirements of the selected data, the determination of the correlation between the data and the content of the project is carried out. For example, according to the first project topic, one of the indicators of data collection, such as the world atlas published by a more famous publisher in foreign countries, or the atlas published by the relevant countries, we can select the file number and related data item information of the foreign atlases in the current collection of National Surveying and Mapping Archives, as shown in table 1.

Tab.1 The selected metadata from foreign atlas in National Surveying and Mapping Archive of China

Archival code	Name of atlas	Publishing unit	Year of publication
TD413GW-L0015-01	Nouvel Atlas Mondial	Beite'ersiman Cartographic Institute	1977
TD413GW-M0034-01	The National Atlas of Japan	Geographical Survey Institute Japan Map Center	1977
TD413GW-M0007-01	THE INTERNATIONAL GEOGRAPHIC ENCYCLOPEDIA AND ATLAS	Columbia University Press	1979
TD413GW-S0010-01	The Pocket Atlas,Japan	Japan Pingfan co.LTD	1980
TD413GW-M0002-01	The Daily Telegraph WORLD ATLAS New Revised Edition	Collins and Son Publishing Company,UK	1980

After these atlases are selected, the map of the project area is carefully read and screened, and the suitable map is selected.

The above methods show that it is more efficient to select the qualified map data through the metadata of existing historical map data.

However, the map archives data selected by this method cannot satisfy the needs of the special topics. For example, in terms of time span, the breadth of the publishing country, and so on, the needs of the project cannot be met. The scope of material collection must be expanded, so target collection is required.

1.2 Target collection

Target collection is divided into two ways: procurement and network acquisition. Procurement is usually a targeted purchase of specific information for an item from a map provider, such as a sea area topographic map published by a country at a certain time. The procurement of map data is generally guaranteed in terms of quality. But in the aspect of metadata acquisition, there is more information to be recorded by manual acquisition, that is, there is more metadata information to be verified. In the way of network acquisition, the network crawler software is usually used for data capture. You need to set up restrictions for different URLs, such as subject management, download time, file type, etc., so as to minimize the scope of filtering and get them on relevant websites through keywords. The correlation of the map data obtained by this method of acquisition is not necessarily guaranteed. But some websites offer better quality maps, such as The World Digital Library, the Library of Congress, and so on. The map information provided by these large libraries can even provide some metadata information, but how to get the metadata information smoothly is not easy.

Some map and picture metadata information can be read out by some image decoders, but there are still problems of incomplete information, which need to be supplemented one by one. Some libraries also provide relevant software to help users download metadata while downloading maps and pictures. However, metadata acquired by this way needs to read its metadata information through special programming, and then match with the designed specific metadata items. The test process may be repeated many times. Many of the metadata information related to the project need to be checked and recorded one by one, and the efficiency of the work is not necessarily high.

2. The principle of metadata design for historical map archives

No matter what kind of historical map data we get, there is no difference in supporting projects. Therefore, the principles to be considered in the design of metadata in historical map are unified, that is, there is more metadata information to be verified.

2.1 Principle of applicability of user requirements

At present, the main goal of the spatial metadata standard developed by the standardization organizations at home and abroad and the spatial metadata standards formulated by other individuals or organizations is the sharing of data. In order to achieve the purpose of sharing, it is necessary to meet the needs of the main users as much as possible in the design. When the project is designed and approved, the users of the project are identified as experts and scholars interested in the relevant issues in the East China Sea and the South China Sea, and all kinds of information resources should be provided as comprehensively as possible to meet the needs of the users. In this project, in addition to the map data, it also involves images, articles, monographs and video, and the design of the corresponding metadata information. Experts should be concerned about the contents expressed in historical map data, the time reflected in historical map data, the mathematical basis of map data, the production units, the language of map data, and the degree of control of data, so the setting of these data items is necessary.

2.2 Standardization principle

It is the most efficient and easiest interoperation to achieve "spatial metadata interoperability" by conforming to the metadata standard. Therefore, it is very important for us to follow the existing standards to achieve interoperability. In this project, the data items that can adopt the relevant standards are all applied to the relevant standards. For example, the format of coordinate information of hot spots in maps and the format of data providing unit coordinates are based on the current common data format of map coordinates. It is beneficial to annotate or obtain the corresponding coordinate point information on the network map such as the Map World. And for time data items, they are required to be unified into the year of the ad.

2.3 Simplification principle

Retrieval resources tend to focus on applicability, while portal applications emphasize simplicity. At present, the metadata items contained in the spatial metadata standards developed by the major standardization organizations at home and abroad are very detailed and complex. The metadata of this project can not only achieve the simplicity of portal application, but also meet the applicability of retrieval resources. For example, the setting of data items of project topic correlation description and key term data items of map data is to facilitate the use of keywords. In this project, the setting of keywords in the South China Sea, the East China Sea and the Diaoyu Dao is to enable users to quickly understand the relevant information of map data and quickly retrieve map data of relevant regions.

2.4 Principle of generality and specificity

The principle of generality of metadata scheme requires that a more general concept can meet the requirements of description, while the principle of specificity refers to the special requirement of resource description in particular domain. At present, the main standardization organizations at home and abroad have established spatial metadata standards, generally speaking, to meet the general data sharing needs of users, the universality is obvious. And in the design of this project, more consideration is given to project topics, which meet the requirements of information acquisition for spatial map and time series as well as publishers and providers, so the specificity is more obvious. In addition, the naming of historical maps has also made specific requirements. The historical map data managed by various units, some were named by file numbers, some by map names, and some by simple serial numbers. In order to facilitate the management, according to the project requirements, the naming rules of the map file are the way of country + time + map.

3 Design of metadata items for historical map archives

For the data collection of this project, in addition to the regular use of library collections, purposeful purchase, and exchange and so on, collecting online through the Internet is also a relatively large source of collection. The location of the relevant information and the information it can express, reflect the position must be determined. Therefore, we cannot use the data file catalog management mode to design metadata items and data content. We should consider more description information and reflect the content characteristics of historical map data more through metadata. In addition, the intellectual property rights of historical maps should be considered, which should reflect the names of individuals and institutions. As a result of the sensitivity of the position of the information, the control of the data should also be considered. In reference to the World Digital Library digital collection description metadata technical standard (http://project.wdl.org/standards/metadata.html, WDL Descriptive Metadata Element Set), electronic document metadata program for clerical purposes DA/T 46—2009 the archives industry standard of the people's Republic of China, the metadata stipulate of the catalog service system of the surveying and mapping geography information achievement, Library of Congress website (http://www.loc.gov), University of Wisconsin Library website (http://www.library.wisc.edu, University of Wisconsin Digital Collections)and so on the design model of some metadata, the data items of the historical map of the project (including the atlas) data metadata are identified, specifically in table 2.

In order to better input and manage the metadata of historical map data, a metadata entry system for historical map data has been developed. The interface of the system is as shown in figure 1. The data can be imported in batches and checked and verified one by one to ensure the quality of data. There are special cues for required entries and thumbnails for data graphics, which can directly check whether the situation of historical map data is normal or not.

Fig.1 The historic map metadata entry interface

Tab.2 The project historic map metadata

Metadata item name	Specific instructions and points for attention	Example	Data type	Is it optional or must be filled?
name	The name of the map is the name of the map, if the name of the map, for historical names, it is necessary to include the corresponding region name, Chinese translation of the name	East Indies (Indonesia, Philippines); Indochina (Vietnam, Laos, Kampuchea)	char	required
electronic document name	Name the picture first, then copy and paste it to the map metadata table in the corresponding cell, not only to ensure that the	(Germany) 1899 An Shi World General Atlas P121-122 Asia. JPG. (China) 2015 Map Publishing House Map of the People's Republic of China.jpg.	char	required

	naming and data names are exactly the same.			
description	A brief description of the features of the map, especially its historical background	General purpose: explain the time, location, author, location, time and reason of the data; the key hints of the information found in the chart. As in this picture, the Diaoyu Dao and Chiwei Yu are represented by digital symbols and footnotes	char	required
author	It means the author of the data, if there are many authors, fill in the first one, and then replace it with [etc.]. Without the author's premise, it is the government department or the publishing institution, the Chinese translation of the name	China Cartographic Publishing House	char	required
publisher	The name of the publishing unit and the Chinese translation	China Cartographic Publishing House	char	required
country of publication	Generally refers to the publishing country, foreign information	China, Germany, etc.	char	required
createdate	Original map has clear record, must fill in, this is the item that relates to time, can be a time point also can be a time period, fill in in the way of AD chronology, YYYY	2017, the years before the Republic of China were converted into the first year of the public, and the era of the Japanese era was converted into the year of the public.	char	optional
inference time	The original map is not recorded, and a time is deduced from the relevant data	Text, for example: 1840 - 1904	char	optional
language	It refers to the language used in the original data and can be filled in for the ambiguous language.	Chinese, German, other	char	required
indicated area	The region expressed by information must include important geographical areas,	East Asia, China The East China Sea, the Ryukyu Islands East Asia, China, Guangdong Province The East China Sea, the Ryukyu Islands,	char	optional

	countries, regions, or important geographical names involved.	Amami Islands		
scale	The main graph of the original map is a scale, such as the original map has no scale, but it cannot be filled in.	For example：1:1 500 000	char	optional
Map attention point coordinate information	Map - related hot - spot coordinates, one or more groups	First latitudes after the longitude, fill in the format as follows:10° 30′ 00″ N,109° 30′ 00″ E	char	optional
Providing unit coordinates	Coordinates of the units that provide information	39° 52′ 37″ N,116° 21′ 15″ E	char	optional
coordinated system	If the original map has, it should be filled; If the original map doesn't have, it shouldn't be filled.	Beijing Geodetic Coordinate System l954	char	optional
vertical datum	If the original map has, it should be filled; If the original map doesn't have, it shouldn't be filled.	Huanghai Vertical Datum 1956	char	optional
map projection	If the original map has, it should be filled; If the original map doesn't have, it shouldn't be filled.	Gauss - g Lv Ge projection	char	optional
the name of the atlas	The name of the collection that information belongs to.	Atlas of the World, Chinese Atlas	char	optional
The name of the provider	The name of a research institution or library or archives.	China Map Publishing House,National Geomatics Center of China	char	optional
URL	If the data is provided by an organization online, the URL link address is provided	https://www.loc.gov/item/gm71005060/	char	optional
Control degree of original graph	The degree of control of the data is controlled in the open publication.	Open, controlled	char	required
ID	8 - bit alphabet and digital information encoding for each piece	DT00001	char	system generation

		of data			
keyword	The key elements that are not present in the above table	Such as the intermittent line, the Diaoyu Dao,the Dongsha Islands, the Zhongsha Islands, Paracel Islands, Spratly Islands	char	optional
remarks	Information that needs to be added to the information		char	optional

4 Complimentary close

The metadata of the special historical map database is different from the metadata in the general database, because its data source metadata is relatively complex. In order to improve the efficiency of the use of historical map data, the design of historical map description information data items can let the users of the data quickly understand the background information of the map, and do not need to extract and synthesize the scattered data item information. For the convenience of query and use of historical map data, the metadata items for spatial positioning of historical map are designed, which greatly improves the efficiency of historical map and will broaden the application field of historical map data.

REFERENCES

[1] Wei Siqi. Object-Oriented Geospatial Information metadata Analysis [J]. Geospatial Information, 2012, 10(01):22-24, 1.

[2] Li Li, Chen Lifang. Construction of Wuhan Historical Map Resource Database [J]. China Archives, 2016(4):56-57

[3] Liu Wenbing. The practice of digital map resource publishing and retrieval based on Web Services [D]. Zhengzhou: Information Engineering University, 2013.

[4] WDL Descriptive Metadata Element   Set [DB OL].http://project.wdl.org/standards/metadata.html

[5] Anonymous name. Document industry standard DA/T46-2009of the People's Republic of China, document electronic file metadata scheme [J]. Science and Technology Archives, 2011 / 3: 39-40.

[6] Zhang Li, Gong Jianya. Research and implementation of Geospatial metadata Management [J]. Journal of Wuhan University of Surveying and Technology. 2000, 25(2):127-131.

[7] Shen Tiyan, Cheng Chengqi. Design and implementation of Geographic metadata Technology system [J]. Journal of Wuhan University of Surveying and Technology.   1999, 24(4):326-330.

[8] Liu Jiazhen, Liao Ru. Quality Control and Management of metadata in Electronic document Management [J]. Library and Information knowledge, 2009(6):91-96,102.

[9] Niu Jinfang, Zheng Xiaohui, Zeng Ting, et al. A metadata Program for the Library of Tsinghua University [J]. Journal of the University Library, 2003(2):22-25.

[10] Zhang Xiaolin, Editor-in-Chief, metadata Research and Application [M] .Beijing: Beijing Library Publishing Press, 2002.

[11] Xu Shaokun, Song Nationalist, Chen Linyu, et al. Research on the application of multidimensional visualization technology in geospatial metadata retrieval [J]. Geomatics World, 2013, 20(1):46-50.