International Cartographic Association
Association Cartographique Internationale

A Electronic Map Data Model Based on PDF

ZHOU Xiaodong,[a,b] YANG Chuncheng,[a,b] MENG Nina,[c] XIE Peng[a,b]

[a] *State Key Laboratory of Geo-Information Engineering, Xi'an 710054, China*
[b] *Xi'an Research Institute of Surveying and Mapping, Xi'an 710054, Chinac*
[c] *Chang'an University, Xi'an 710054, China*

Abstract: In this paper, we proposed the PDFEMAP (PDF electronic map) that is a kind of new electronic map products aiming at the current situation and demand of the use of electronic map products. Firstly gives the definition and characteristics of PDFEMAP, followed by a detailed description of the data model and method for generating PDFEMAP, and finally expounds application modes of the PDFEMAP which feasibility and effectiveness are verified.

Keywords: PDF, Electronic Map, PDF Maps, Digital Map

1. Introduction

Since its birth and development, the map has become an indispensable and important tool to know, understand and utilize the geographical environment. It has been widely used in economic, social, cultural, military and other industries. Maps have scientific value, social value, legal value, cultural value and military value, which importance for the evaluation of human society can not be overemphasized [1].

In the information age, with the changes of the analog map to the digital map, the constant emergence of a variety of special map products based digital map and information systems which providing all kinds of map service, makes the target of map users original professionals transform to public user [2,3].

Electronic map is becoming a popular information product, and will serve as an important tool for mass communication [4]. However, the current electronic map production and application require specialized electronic map software. Map services are inseparable with the corresponding geographic information system. Which makes very valuable geospatial information is closed in the complex application software, and only some professionals know how to use [5,6,7].

In general, non-professional users also need to share geospatial information. More importantly, maps in these systems use way of graphical representation. This visualization is based on geographic information, which can quickly access the symbolic map through the symbol mapping mechanism, and can maintain the correspondence with the geographic information before symbolization. But the graphical representation can't get standard map in line with the publication effect [8,9].

The current published map is usually edited and completed in the professional mapping software. That is the issue of integration of maps and geographic information, on which the current scientific community and the industry having no good solution. However, there is a growing demand for standard maps by GIS users, who want to see the map beautifully when reading, rather than looking at rough maps that affect people's perception and decision-making. Therefore, the user needs a map product containing both geographic information data and the corresponding map symbol data, which takes into account the processing needs of the computer and human reading experience.

PDF (Portable Document Format) is a document recording format which developed and designed by Adobe company. Containing all of the data needed to display and print out on a printer, image setter or Computer-to-Plate, and completely independent of the original application and computer platform from which it was created, PDF document already become ideal for electronic document transmission and digital dissemination format.

At present, China has gradually begun to the research of storing maps and geographic information by using PDF documents. On which there are some relevant research results [10,11]. In the deep study of PDF electronic map (PDFEMAP), the definition of PDFEMAP is given, and the characteristics of PDFEMAP are extracted. In particular, PDFEMAP data model and application pattern are explored in a meaningful way.

2. PDFEMAP Definitions and Characteristics

PDFEMAP is a new type of electronic map, which contains both geographic information and symbolic graphics. Based on the PDF 1.7 specification, it conforms to the ISO 32000-1 format standard [12-15]. It has geographic coordinates query, hierarchical browsing of geographical features, geospatial object query, direction calculation, length and area calculation, information annotation and other functions.

PDFEMAP enables users, both professional and layman, to read hierarchically geo-referenced maps and imagery products using PDF reading tools such as Adobe Acrobat, Adobe Reader, Foxit Reader, and so on.

PDFEMAP product features summed up in the following four points:

1) Product Independence: Being the independent document of PDF form, needing not relate the extra document, this product can be divorced from the GIS software or other complicated visualization system, and be browsed with only any PDF reader;
2) Data Integrity: This product embeds the spatial data and attribute data in the geographical target symbolic graphic data (including images, text, etc.) without associating the map symbol library, which is the integrated whole data;
3) Cross-platform and Cross-terminal Applications: This product based on PDF format, compared with the traditional format of the electronic map, not only has nothing to do with the system platform, but also has small amount of data size, Which can be used in portable terminals;
4) Industry Applicability: Directly to the application, this product does not require GIS-related professional knowledge and complex system training, with which non-professional users can quickly use and share map results.

3. PDFEMAP Data Model

The object of PDFEMAP storage and management is the electronic map, including geometric information, attribute information and symbolic graphics information of the geographic features in three areas. In this paper, the structure of the PDFEMAP storage and management is constructed by combining the object-oriented technology and the PDF technology, thus having the ability to store the electronic map in a PDF file using an object-oriented method.

3.1 PDFEMAP Document Structure

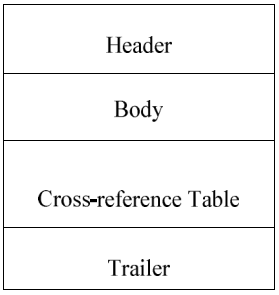

Fig.1. PDFEMAP document composition

PDFEMAP following the PDF format standard, as general PDF document, is composed of four parts as file header, file body, cross-reference table and the end of the file, as shown in Fig.1.

The file header indicates the version number of the PDF specification that the file follows and appears in the first line of the PDF document. PDFEMAP version number is usually 1.6, the file header corresponding to "% PDF-1.6". It should be set to 1.7 if it needs to support the three-dimensional map.

The file body is made up of a series of PDF indirect objects. Each object's hierarchical composition constitutes the PDFEMAP document structure, which is the main part of the PDFEMAP document.

A cross reference table, starting with the 'xref' keyword, is an address index table of an indirect object which is set up to allow random access to an indirect object.

The end of the file is to enable the application to quickly find the cross-reference table and the specific object which start with 'trailer' keyword while reading PDFEMAP document.

PDFEMAP document structure, showing a tree structure, looks the directory object (Catalog) as the root node. The directory object node mainly contains three sub-nodes, namely page cluster (Pages), feature property tree (StructTreeRoot) and layer property tree (OCProperties), as shown in Fig.2.

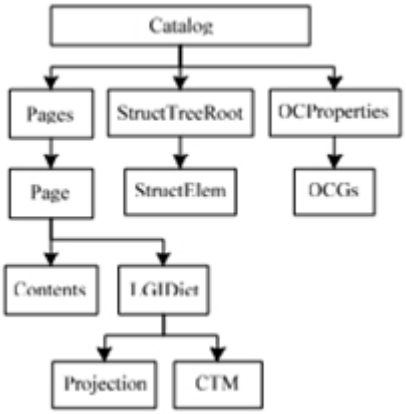

Fig.2. PDFEMAP document structure

A page cluster contains a page object node. Page is the most important component of PDFEMAP, which mainly contains the two object nodes of content flow (Contents) and spatial reference (LGIDict). The content stream is used to store PDFEMAP graphical information, including a series of instructions to draw graphic elements onto the page. The spatial reference is used to convert PDFEMAP page coordinates to geographic coordinates, including the Projection and Transformation Matrix (CTM). Map projection to achieve the transformation between the geographical coordinates and the projection coordinates, which mainly store projection type, projection parameters, ellipsoid parameters and other information. The transformation matrix realizes the conversion between the projection coordinates and the PDFEMAP page coordinates.

3.2 PDFEMAP Feature Organization

PDFEMAP is usually stored in map format. Each PDF file contains a page, storing a digital map. PDFEMAP features organization is organized hierarchically by geographical features which reference digital map data organization. Geographical features include natural geographical elements and social geographic elements, which are the collection of geographical entities with the same semantic characteristics.

The geographic entity is the main content of PDFEMAP, which consists of geometric location information, target attribute information and symbol graphics information. The geometric positioning information and the target attribute information are DLG data, that is, geographical

Proceedings of the International Cartographic Association, 1, 2017. This contribution underwent single-blind peer review based on submitted abstracts | https://doi.org/10.5194/ica-proc-1-129-2017 |

information database data, which are important components of PDFEMAP for querying information. The symbol graphics information is the map expression data of the features, that is, the result data of cartography. The symbol graphics information consists of the basic primitives, and does not depend on the specific technology and symbol library. The symbol graphics information constitutes the map expression effect of the PDFEMAP.

PDFEMAP contains a number of geographical feature layers, which is usually divided into natural geographical feature layers such as physiognomy, hydrographic net, vegetation, soil and so on and socio-economic feature layers such as residential area, road, industrial and agricultural cultural facilities, administrative divisions etc. Each feature contains several geographic entities, usually divided into point-like entities, linear entities, planar entities and bodylike entities. The geographic entity contains geometric location information, object attribute information, and symbol graphics information, as shown in Fig.3.

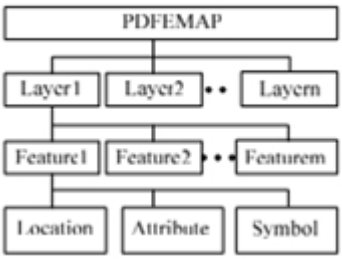

Fig.3. PDFEMAP Element Organization

3.3 PDFEMAP Conversion Output

PDFEMAP conversion output includes three parts of spatial reference system establishment, layer information configuration and geographic entity information output.

The establishment of spatial reference system is to form a map frame, which mainly includes parameter setting of transformation matrix and map projection. The PDF standard defines a device-independent coordinate system, also known as user space, which keeps the relationship of the object on the page constant and thus has a uniform effect on any output device. The PDF user space adopts the rectangular plane coordinate system and realizes the coordinate transformation between the PDF user space and the map plane rectangular coordinate system through the transformation matrix. Transformation matrix with 3×3 matrix, needs to record six parameters [a b c d e f]. Transformation expression as shown in equation (1).

(1)

Where (x', y') is the rectangular plane coordinate of the map and (x, y) is the PDF user coordinate of the map.

The map projection parameters include the ellipsoid parameters and projection parameters needed to define the rectangular plane coordinate system of the map, and coordinate transformation between rectangular plane coordinate system and geographic coordinate system. Assuming that the geographic coordinate of the rectangular plane coordinate point (x', y') of the map is (,) the inverse transformation equation of projection is shown in equation (2).

(2)

The transformation between the PDF page coordinate and the geographic coordinate can be realized by simultaneous equation (1) and equation (2).

First, the layer resource (OCProperties) attribute values, which including layer definition (field as name, attribute,

etc.), layer order, display switch and other information, should be set according to the map features hierarchy information in order to configurate layer information. As the PDF specification does not support GB2312 encoding,

when the layer name and other information is Chinese characters, it should be converted to Unicode encoding.

The symbol graphics information in geographic entity information exports to content stream of the PDF page usually by a series of graphical drawing instructions. This instructions are used to draw graphical information on a PDF page. The drawing order on a PDF page is decided by the graphical drawing instructions. In the same place, the later drawing graph will gland the earlier drawing graph.

In addition, in order to achieve the best effect of the map surface, the gland relationship processing between geographic features in process of cartography is an indispensable important processing link.

After symbolized of the geographic entity, it is necessary to solve the gland problem of symbols. Here, it is proposed that the unit symbol is the smallest unit. The entity symbol usually contains a plurality of unit symbols which priority is often set to be between 0 and 255.

All leaf node objects (StructElem) are enumerated by the keyword "K" when attribute information in the geographic entity information been output to the feature attribute tree. A leaf node object can be used to store a set of custom property data objects (UserProperties) by the keyword "K" to enumerate unique identifier (MCID) which is to associate tagged graphics content.

4. PDFEMAP Application Patterns

PDFEMAP application can be used in two patterns: First, open PDFEMAP data file using PDF reader, and use

PDFEMAP function; Second, using hypermedia technology, the establishment of electronic map system can be based on Windows, Android, IOS and other operating systems. The first model is simple, easy to control, and is more suitable for non-professional users. The second model based on PDFEMAP, adopting hypermedia technology, considers geospatial entity as a "node" and relationship between the geographical entities as "Chain". By using the interrelationship between the node and the chain to build complex information network, the PDFEMAP would be with beautiful graphics, simple human-computer interaction and other characteristics, and suit for various types of users.

Frequently-used PDF reader such as Adobe Acrobat, Adobe Reader, Foxit Reader, etc., after installing the

Proceedings of the International Cartographic Association, 1, 2017. This contribution underwent single-blind peer review based on submitted abstracts | https://doi.org/10.5194/ica-proc-1-129-2017 |

official software, which can enable the users to achieve the PDFEMAP reading and browsing, layer control and geographic object capture and other functions. If additional functions such as spatial location information query, geographic measurement, feature attribute query and some simple GIS spatial analysis function is needed, the PDFEMAP plug-ins (Plug-ins) corresponding to the PDF reader should be installed.

Utilizing the secondary plug-in development capabilities provided by Adobe Acrobat, the developer can develop PDFEMAP plug-ins with the functions of map query, map measurement etc., and realize PDFEMAP browsing, query and measurement.

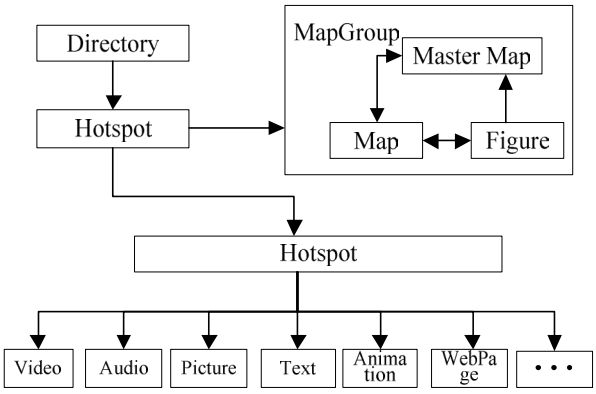

Fig.4. PDFEMAP hypermedia applications

Hypermedia is composed of nodes that store multimedia information and chains that describe the relationship between multimedia information. PDFEMAP based on hypermedia technology is a kind of storage and management of geographic or non-geographic information technology. Information management in accordance with mutual relations is scalable, with direct access through the relationship. Having both aesthetic and spatial analysis capabilities, PDFEMAP provides users with a variety of convenient communication tools such as buttons, hotspot, anchors, and more. In addition to the node type of the general hypermedia system, hypermedia PDFEMAP also contains a lot of nodes for geographical information expression, such as directory, map group, master map, sheet, table and so on, as shown in Fig.4.

5. Conclusion

A PDFEMAP not only has the paper map visual expression effect, but also has some functions of the vector map, can map hierarchical display, target attribute query, spatial coordinate query and spatial measurement and other operations.

As the use of PDF format to store geographic information data, PDFEMAP product terminals do not need to rely on dedicated and expensive paid GIS software, which reduce the user requirements to use PDFEMAP. Setting up a "bridge" between the users of the mapping of professional and non-professional, maximumly closing the distance between the map producer and the user, PDFEMAP satisfies the requirement of using geographic information.

As a new electronic map, PDFEMAP will break the boundary of GIS professional application, and it is of great

significance for non-professional of GIS to use geographic information and improve emergency mapping guarantee

capability.

6. References

WANG Jiayao, CHENG Yi. Discussions on the Attributes of Cartography and the Value of Map[J]. Acta Geodaetica et Cartographica Sinica, 2015,44(3):237-241.

WANG Jiayao. Development Trends of Cartography and Geographic Information Engineering[J]. Acta Geodaetica et Cartographica Sinica, 2010,39(2):115-119.

GAO Jun. Cartographic Tetrahedron: Explanation of Cartography in the Digital Era[J]. Acta Geodaetica et Cartographica Sinica, 2004,33(1):6-11.

WANG Hong, AI Tinghua, ZHU Guorui. Adaptive Strategy on the Visualization of Electronic Map[J]. Geomatics and Information Science of Wuhan University, 2004,29(6):525-528.

AN Yun. Electronic Map VS Paper Map [J]. MAP, 2006(03):26-27.

YANG Chuncheng. Electronic Map Making and Application System Design Under WIN32 [J]. Journal of Institute of Surveying and Mapping, 2001(03):220-222.

LIU Haiyan, HOU Suyuan. The Research for Using DirectX in the Electronic Map [J]. Journal of Institute of Surveying and Mapping, 2001(01):50-54.

LI Lin, ZHU Haihong, HE Biao, etc. Cartographic Model for Topographic Maps Based on Algebraic Structure [J]. Acta Geodaetica et Cartographica Sinica, 2011,40(3):373-378.

LI Lin, XU Ming, YIN Zhangcai, etc. Development of map based visualization of geographical information[J]. Engineering of Surveying and Mapping, 2006,15(4):11-14.

HU Haojie, WEI Bin, YANG Chuncheng, etc. Method of Batch Outputing PDF Map Labeling[J]. Journal of Geomatics Science and Technology, 2011, 28(5):378-382.

HU Haojie, YANG Chuncheng, WEI Bin, etc. Method of Outputing High-Quality PDF Map Lattering[J]. Journal of Geomatics Science and Technology, 2010,27(6):450-454.

OGC. GeoPDF Encoding best Practice Version 2.2, O. G. Consortium, 2009.

http://www.acrobatusers.com/articles/adobe-acrobat-and-reader-now-support-geospatial-data-sharing

Raymondgcaputo. The GeoPDF Project: Creating Maps for Non-Mapper[J]. Engineer, 2010, 1:36-37.

http://www.adobe.com/devnet/acrobat/pdfs/PDF32000_2008.pdf

Proceedings of the International Cartographic Association, 1, 2017. This contribution underwent single-blind peer review based on submitted abstracts | https://doi.org/10.5194/ica-proc-1-129-2017 |