

# **Publishing in PDF--A Promising Net Map Publishing Mode**

Liu Zhen, Deng Shujun

Department of Cartography  
Zhengzhou Institute of Surveying and Mapping  
Zhengzhou, 450052, China

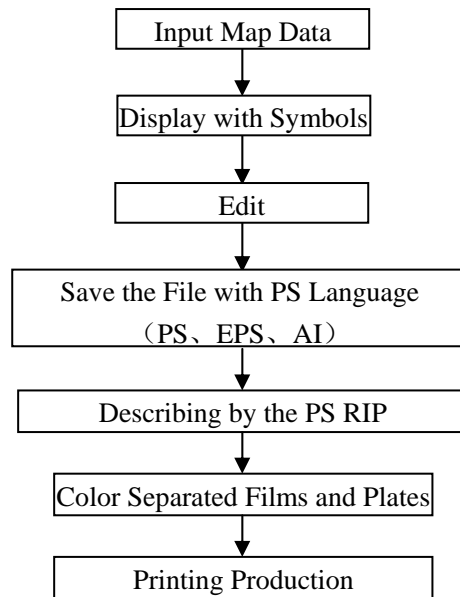
**Abstract:** PDF, Potable Document Format, is an excellent file format based on Postscript language. It has many properties that are compatible with net publishing. Publishing map on net in PDF is undeveloped and promising. This paper discusses several foundational problems on publishing map in PDF.

**Key Word:** PDF, Net Publishing, Postscript Language.

## **1.Foreword**

The development of the digital technology since 1980s and the network technology since 1990s has pushed the digital map making and map publishing forward rapidly. At present, the map making and the map publishing technology are undergoing a transition from the analog production mode on the base of "object carriers transfer + storehouse + transportation "to the whole digital production mode on the base of " digital information treatment + database + net communication ". With the completion of the transition, map making and map publishing, the two different scientific fields, will gradually melt into each other. In addition, the network technology will play an important role in map making and publishing technology of the Net era.

In the early 1980s, appeared a new Electronic Publishing System (EPS) based on PostScript (PS) Language and it soon became dominant in the publishing industry. Seizing the opportunity, map publishing use this technology too. Now the Map Publishing System (MPS) (for example MicroStation, MapGIS ,FangZheng ZhiHui and so on ) on the base of PS Language are widely used in map publishing organizations. The workflow of the systems is as follows.



**Figure1 Workflow of the Map Publishing Mode Based on PS Language**

The map electronic publishing mode on the base of PS Language makes map making and publishing integration, convenient the map production and enhances the quality of the map publishing, so it has the wide use prospect. But most of the EPS are based on the second version PostScript Language, and they have following shortage: First, there is language different between the PS Language used in different System, just like localism. When transported from one system to another, it will go wrong. Second, it has no screen display function. If wanted to see the result, we must print it out. The last, it has no network function.

The PDF map-publishing mode recommend in the paper is also belonging to the map-publishing mode based on PS Language. But PDF is come from PS Language and more advanced than PS Language. Sometimes it is known as based on third version PS Language. The authors think Publishing in PDF--A Promising Net Map Publishing Mode.

## 2.The Future Map Publishing

### 2.1 The Integration of Map Making and Map Publishing

On the basis of the computer science, the technologies of map making and electronic publishing will be melted together organically. As a result, lots of work and

times will be saved, and the quality will be promoted at the same time. The mode of future map publishing is that the whole system is an integrated system included the map design, map edit, individual pages assembling, and output, so it will be more simplified and intelligent.

## **2.2 The Function of the Net Publishing**

"The 19<sup>th</sup> century is the era of the railroad, the 20<sup>th</sup> is the era of the High way, and the 21<sup>st</sup> century is the era of the information super high way." Now we have come into the era of the Internet and the World Wide Web. In this era, with the limitation of the Desk Top Publishing mode broken, map publishing will enter a net publishing environment. The future map publishing system should not only the function of real time net transferring, but also the function of distant controlled output.

## **2.3 An Open Way of Map Publishing**

"Open" is one of the words that we come across the most often when taking about the information technology. It includes the open interface, the open data structure as well as the compatibility used in different computer platform. At present the electronic map publishing system being used is not open enough. To acquire the satisfactory effect we have to use Intergraphy publishing software to publish the maps made with Intergraphy software, or we have to use FangZheng publishing software to publish the maps made with FangZheng software. Unexpected errors will occur if different company software are used, and this will be the key problem we should solve during the net publishing.

In the net publishing environment, what we concern about is not the sources of maps or the software used in making the maps, but the map last products we needed. Thus all the mapmaker should follow a certain standard when publishing maps. PDF is an excellent standard, which has been widely accepted in electronic publishing field. It should also be adopted in map publishing.

## **2.4 "What We Get is What We See" Map Publishing**

That "what we get is what we see" is a very old subject in the print industry. It remains to solve at present. We often meet with the problem that what is printed is not the same with what we see on the screen, especially the color and character style. To

solve the problem involves the hardware, software as well as our visual characteristics. It is almost impossible to achieve the perfect result that what we get is the exactly same as what we see. PDF files contain a lot of information (e.g. color management information and font information). Publishing maps with PDF can meet the requirement to a great extent that what we get is what we see.

### 3. Publishing in PDF--A Promising Net Map Publishing Mode

#### 3.1 What is the Portable Document Format?

PDF is a file format used to represent a document in a manner independent of the application software, hardware, and operating system used to create it. A *PDF file* contains a *PDF document* and other supporting data.

A PDF document contains one or more pages. Each page in the document may contain any combination of text, graphics, and images in a device- and resolution-independent format. This is the *page description*. A PDF document may also contain information possible only in an electronic representation, such as hypertext links, sound, and movies.

In addition to a document, a PDF file contains the version of the PDF specification used in the file and information about the location of important structures in the file.

#### 3.2 The Foundational Properties of PDF

Given the goals and intended use of PDF, its design has several notable properties. This section describes those properties.

##### 3.2.1 Adobe imaging model

PDF represents text and graphics using the Adobe imaging model; this is the same imaging model that is used by the PostScript language. Like a PostScript language program, a PDF page description draws a page by placing “paint” on selected areas.

##### 3.2.2 Portability

A PDF file is a binary file; the entire 8-bit range of characters may be used.

Unfortunately, some agents treat files that happen to use only the printable subset

of the 7-bit ASCII code and white space characters as “text,” and take unreasonable liberties with the contents. For example, mail-transmission systems may not preserve certain 7-bit characters and may change line endings. This can cause damage to PDF files.

Therefore, in situations where it is possible to label PDF files as “binary,” we recommend that this be done. One method for encouraging such treatment is to include a few binary characters (codes greater than 127) in a comment near the beginning of the file, *even if the rest of the file is ASCII*. This ensures that a PDF file will be treated as binary when this is possible, while still allowing it to be transferred through a non-binary channel without damage.

### **3.2.3 Compression**

To reduce file size, PDF supports a number of industry-standard compression filters:

- JPEG compression of color and grayscale images.
- CCITT Group 3, CCITT Group 4, LZW (Lempel-Ziv-Welch), and Run Length compression of monochrome images.
- LZW and Flate compression of text, graphics, and indexed image data.

### **3.2.4 Font independence**

Managing fonts is a fundamental challenge in document exchange. Generally, the receiver of a document must have the same fonts the sender used to create the document. Otherwise, a default font is substituted, producing unexpected and undesirable effects because the default font has different character metrics (widths) than the intended font. Differing font-widths may cause lines to extend into margins or overlap graphics.

The sender could include the fonts with the document, but this can easily make even a short document quite large—a typical two-page memo using four fonts might grow from 10K to 250K. Another possibility is that the sender could convert each page of the document to a fixed-resolution image like a facsimile. Even when compressed, however, the image of a single page can be quite large (45–60K when sampled at 200 dpi). In addition, there is no intelligence left in the file, preventing the

receiver from searching for or extracting text from the document.

PDF provides a new solution that makes a document independent of the fonts used to create it. A PDF file contains a *font descriptor* for each font used in a document. The font descriptor includes the font name, character metrics, and style information. This is the information needed to simulate missing fonts and is typically only 1–2K per font. If a font used in a document is available on the computer where the document is viewed, it is used. If it is not available, a multiple master font is used to simulate on a character-by-character basis the weight and width of the original font, to maintain the overall “color” and formatting of the document. This solution applies to both Adobe Type 1 fonts and fonts in the TrueType™ format developed by Apple Computer, Inc.

### **3.2.5 Random access**

Every PDF file contains a cross-reference table that can be used to locate and directly access pages and other important objects in the file. The location of the cross-reference table is stored at the end of the file, allowing applications that produce PDF files in a single pass to store it easily and allowing applications that read PDF files to locate it easily. Using the cross-reference table, the time needed to view a page in a PDF file can be nearly independent of the total number of pages in the document.

### **3.2.6 Incremental update**

Applications may allow users to modify PDF documents, which can contain hundreds of pages or more. Users should not have to wait for the entire file to be rewritten each time modifications to the document are saved. PDF allows modifications to be appended to a file, leaving the original data intact. The addendum appended when a file is incrementally updated contains only the objects that were modified or added, and includes an update to the cross-reference table. Support for incremental update allows an application to save modifications to a PDF document in an amount of time proportional to the size of the modification instead of the size of the file. In addition, because the original contents of the file are still present in the file, it is possible to undo saved changes by deleting one or more addenda.

### **3.2.7 Extensibility**

PDF is designed to be extensible. Undoubtedly, developers will want to add features to PDF that have not yet been implemented or thought of. The design of PDF is such that not only can new features be added, but also applications that understand earlier versions of the format will not completely break when they encounter features that they do not implement.

### **3.3 The Significance of Publishing Maps with PDF**

Now, The electronic publishing products described in PDF are widely used in every field. PDF file is well known by more and more people, so publishing map in PDF is of great significant.

#### **3.3.1 Making the Map Publishing More Convenient and Reliable**

Publishing the print products from a map data is the important use of map data. From the point of publishing function, we can divide the Digital Map Making System (DMMS) and Geographic Information System (GIS) into following categories.

- Having no publishing function. The result of the system is map data or electronic map. If printed products needed, the result should be transformed into other systems.
- Having partly publishing function. The system have laser or ink-jet print function, or it can transform the electronic map into some imagine format (e.g. BMP, GIF, JPEG). But this is not the publishing in the usual sense, and the result is not satisfactory.
- Having professional publishing function. The system can transform the electronic map into the file form based on PS Language (e.g. PS, EPS, AI), and output the color separated films by EPS, and then printed the output into printing products.

At present, most of the DMMS and GIS (especially the systems which developed domestic) have no professional publishing function. The system having the publishing function (e.g. Microstation, ArcInfo, MapGIS) cannot transform the digital map into PDF, but PS or EPS files. The reason is the Raster Imagine Processor (RIP) supported

PDF appeared only in recent couple years.

The advance of PDF is:

- Enable the system to have professional publishing function and support the net publishing.
- Using the flexible and portable characteristic of PDF efficiently, to show the map on screen in the "what we get is what we see" mode.
- Fitting the Map Publishing to Electronic Publishing Industry. Because publishing in PDF Format is dominant in Electronic Publishing Industry.

### **3.3.2 Widen the Use of the Digital Map in the Net Environment**

Net Work is the most important method to transmit the information recently. If someone keeps his/her eyes open on the web you can find that a lot of electronic publishing products on the web are described by PDF. When you have got such file, you can read and print them with Acrobat Reader, and also you can put them into other system to edit and revise them. If the electronic map format is GIF or JPEG format, which is widely used on the web, it will lose the vector characteristic. In this case, the use of the map will be limited, even include the simple manipulation, for example, zoom in and zoom out. Transform the digital map into PDF file that has the vector characteristic will increase information of the web map and widen its use area.

### **3.3.3 Contenting the Requirements of the Future Map publishing in High Lever**

PDF is developed by Adobe that is second to none company of the publishing industry. From the beginning, the requirements of web publishing are involved in the consideration. It is from the Desk Top Publishing (DTP) and advanced than the DTP. The publishing with PDF can overcome almost all shortage that publishing with PS or other methods. It can contents the requirements of the future publishing in high lever.

## **4. What Should We Do to Publishing Map in PDF**

### **4.1 The Theoretical Study of Publishing Map in PDF**

The theory is the precursor of the technology, and the technology with no theory supporting is blind. There is application of publishing the common pages in PDF in the developed countries,



and there are some corresponding technology sources. But it is an undeveloped field not only in foreign but in the domestic to publish map in PDF. The theory we should study includes: PDF publishing theory, map cartography and publishing theory, how to combine PDF and map, and the model and workflow of using PDF to map network publishing.

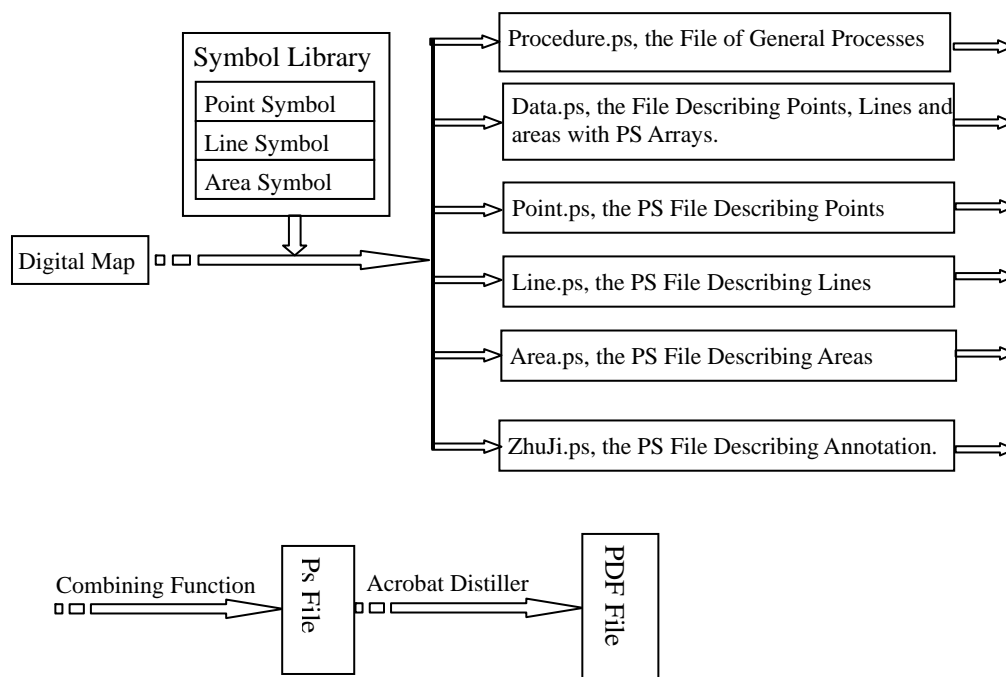
## 4.2 Saving Digital Map As PDF

The precondition of publishing map in PDF is saving digital map as PDF. There are two basic ways, direct and indirect pattern.

### 4.2.1 Indirect Pattern

Indirect pattern is that producing the PostScript Pages describing digital map first, transforming the PostScript file to PDF file with Acrobat Distiller (the Adobe company provide the software for free which are specially used to produce PDF file) then. The center is producing the PostScript pages. The method is easy to understand and master, and the quality of the PDF file is higher. But it is only a transitive method, and don't agree with the request of map cartography and publishing integration.

The authors used the indirect pattern to transform a relief map which scale is 1:250000 to PDF file. The specific steps are described Figure 2.



**Figure 2 The General Workflow Building PDF of The Authors**

- 1 To set up the PS symbol library that describes the point, line and area. The PS symbol library is series of processes of PostScript language, and the coordinate of the point, line and area is the formal preference.
- 2 To Link the PS symbol library to digital map, and produce all sorts of temporaries PS files (Procedure.ps, Data.ps, Point.ps, Line.ps, Area.ps, zhuJi.ps).
- 3 Taking account of the cover relationship of the features and the simplicity of the PS documents, Organize all sorts of temporaries which are already produced to a complete PS file.

4 Transform the PS file that is already produced to PDF file with Acrobat Distiller. In this step, page setting, compress choice and encapsulating font are considered.

#### 4.2.2 Direct Pattern

As the introduction before, indirect pattern is only a transitive pattern. The really productive style of PDF file describing the map should be that the map is saved as PDF directly in the surround of map cartography. Its process is displayed in the figure 3.

Direct pattern is an ideal pattern that can save the digital map as PDF. But it is more difficult to realize. It request that we understand the structure of PDF further, and we also should do some studying about the describe pattern of the map symbol in the PDF file.

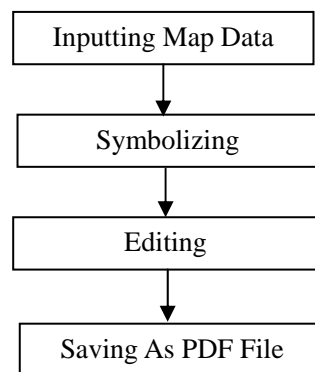


Figure 3 The Workflow Of Direct Pattern

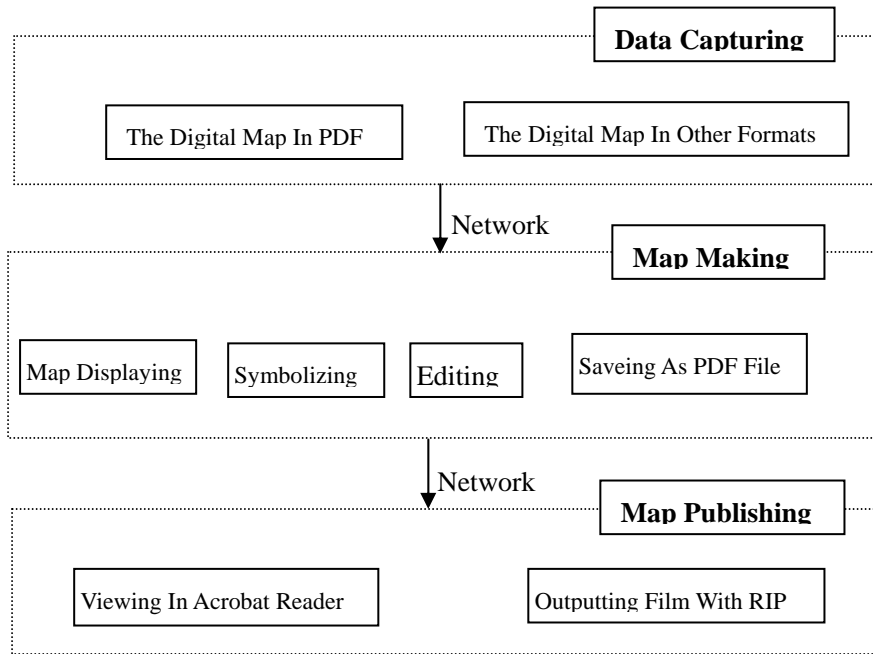
#### 4.3 The Research of RIP supporting PDF and contenting the particularity of map publishing

RIP is a key technology of electronic publishing system. The open degree, output quality and output efficiency of the electronic publishing system greatly depends on the quality of RIP. The principle on which RIP works is extracting the page elements such as graph, image, text from the file that describes the page, transforming these page elements to raster data, then sending these raster data to output device and producing the film.

RIP that supports PDF is necessary to use PDF to publish map. In recent two years, various RIPs supporting PDF have been produced, including Script Work<sup>®</sup> of Harlequin Inc, Apogee PDF of Agfa<sup>®</sup>, PSPNT2.0 RIP of EFounder<sup>®</sup> etc. We can publish map with these general RIP. But the efficiency and quality may be not perfect because of ignorance of the particularity of map publishing such as map symbol base, screen line etc. So developing the RIP supporting PDF and contenting the particularity of map publishing is significative.

#### 4.4 The Development of the Software Integrating Map Net Transmitting, Map Making and Map Publishing

It is necessary for Map Net publishing in PDF to develop Series of Software and hardware, especially a set of software integrating map transmitting, map making and map publishing. The software includes three modules, such as data capturing, map making and map publishing. The modules are connected each other through, that is represented in figure 4.



**Figure 4 The Form Of Software**

## Reference

- [1] **Portable Document Format Reference** Version1.3 *Manual* Adobe Systems Incorporated March 11, 1999
- [2] **PostScript language reference manua** third edition Adobe Systems Incorporated February 1999
- [3] **PostScript Language Document Structuring Conventions Specification** Version 3.0 Adobe Developer Support 25 September 1992
- [4] **Encapsulated PostScript File Format Specification** Version 3.0 Adobe Developer Support 1 May 1992
- [5] Qin Yixin **PDF File and Its Application in Map Electrical Publishing** Zhengzhou Institute of Surveying and Mapping April, 2000
- [6] Jin Yang **PDF and Its Effect on Preprint Disposal** print technology 1999(6)