Current status and future development of Chinese Government Public Information Online

LIANG Huiwei & WANG Zhigeng
Digital Resource and Service Department, National Library of China, 100081 Beijing, China

Abstract Since its launch last year, the Chinese Government Public Information Online (CGPIO) platform’s basic construction has developed rapidly. This paper describes the technology and service status of the platform, analyzes its problems, and details the future development of the alliance platform in the future.

Keywords Government information, National Library of China, Chinese Government Public Information Online, National public library union for integration and services of government information.

1 Introduction
With the development and implementation of the Regulation of the People’s Republic of China on the Disclosure of Government Information (hereinafter referred to as the Regulations)\(^1\), China has gone through many changes in the information disclosure and directory construction of government information. To provide the public with a convenient and practical way in which to access the government’s public information, several factors must be considered including the amount of information to be disseminated, content integration and organization, and the provision of a unified and open information services platform.

As an important national knowledge infrastructure, the National Library of China (NLC) has been working to promote public culture resource sharing in China, which is important for public information access. As a professional institution in information organization and management, NLC is also responsible for the public’s access to government information. As such, NLC launched China’s first integrated information platform, called the Chinese Government Public Information Online (CGPIO) on April 30, 2009 (http://govinfo.nlc.gov.cn/). As well as being an innovative data service, the Platform is the first to provide an integrated government information service by a library.
2 Platform overview

CGPIO aims to comprehensively collect and integrate information from various levels of government into a freely available public service and information portal. The Platform is not only a convenient way for the public to get the government public information, it also helps provide basic resources and services to various libraries, which enhances their role as public information depositaries, integrators, and communicators. The overall framework of CGPIO is shown in Fig. 1.

![Diagram of CGPIO framework](image)

Fig. 1 The framework of the CGPIO.
At present, the information in the Platform is collected by the crawler or similar technologies automatically. However, we also considered to start collecting information under cooperation with government agencies. After intelligent processing of information such as filtering, extraction, and automatic classification, the collected information of the Platform is then classified into three categories: public information, gazette, and government agencies. Furthermore, print information and electronic information held by NLC are also linked with the Platform to provide further services to public users, such as information browsing, information retrieval, and delivery of print copies of government information.

Currently, public information from hundreds of government websites including central government, provincial government, provincial cities and municipalities has been collected, processed and classified into above three categories. The information collected amounts to more than 400,000 items, spanning more than ten years, simultaneously, all of which have been integrated with the NLC’s collection. The category of government agencies includes more than 3,000 organizations. The home page of the Platform is shown in Fig. 2.

![Fig. 2 The home page of the Platform.](image-url)
3 Platform infrastructure and modules

When designing an targeted public platform for government information, the principles of openness, scalability, economics, and security must be followed. It is also important that it be reliable, stable, accurate and timely. The logic infrastructure is required and the advanced technologies should be involved, so that it is easy to encompass new functions and satisfy the current data processing requests and future demands. The system infrastructure is shown in Fig. 3.

The finished function modules in the application layers are listed below:

- **Web information harvesting**  The module tracks and monitors the public governmental information in order to finish incremental harvesting. At the same time, it supplies the harvesting, controlling, and managing functions for the administrators. It also finishes the data processing cooperation with the data processing module.
- **Data processing**  The module automatically finishes data processing based on preset information classification rules, information identification and extraction rules, and information indexing and conversion rules. The processed data is then put into the database.
- **Information release**  The module manages, modifies, and releases the data in the databases. The portal style based on the templates is supported, so are the interactive features, such as the category navigation, multi-level browsing, subscription pushing.
- **Resource retrieval**  The module provides the simple, advanced, and cross retrieval of the government information.
- **System management**  The module offers the administrative staffs the authority control, user management, logs, etc.
- **Accessing and statistics**  The module offers the statistics reports and results such as access and service statistics.

In the presentation layer, two main services are supplied, information browsing and information retrieving. The former is to provide users with public navigation for various types of information. The latter is to assist users to find government public information by various retrievals.

4 Service status

To optimize the Platform’s service, we conducted several statistical tests in regards to one-year use (Table 1). The following indicators were used:

- **Page views**  The actual number of pages clicked, it is often used to measure the popularity of web content and visits.
Fig. 3  The system infrastructure.

Note: CDS: Content distribution server; CKM: Chinese knowledge management, an implementation of low level mining algorithm to provide stable and encapsulated APIs for the Chinese text mining application; Info Radar: customize the information sources of intending to be acquisition and store the information locally to guarantee the information is read time and comprehensive.
• Number of unique visitors The IP numbers that access the site.
• Net viscosity The average visit time by all users within a specified period.
• The number of returning visitors
• Depth of visit The number of pages requested each time.

Table 1 Usage statistics

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Page views (page)</th>
<th>Number of unique visitors (person)</th>
<th>Net viscosity (minutes)</th>
<th>Number of returning visitors (person)</th>
<th>Depth of visit (≥20pages)</th>
</tr>
</thead>
<tbody>
<tr>
<td>May 2009</td>
<td>25,761</td>
<td>1,953</td>
<td>12’2”</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>June 2009</td>
<td>138,737</td>
<td>7,314</td>
<td>18’15”</td>
<td>6,223</td>
<td>40.11%</td>
</tr>
<tr>
<td>July 2009</td>
<td>698,516</td>
<td>8,367</td>
<td>16’34”</td>
<td>4,271</td>
<td>39.08%</td>
</tr>
<tr>
<td>August 2009</td>
<td>1,438,326</td>
<td>9,700</td>
<td>18’49”</td>
<td>4,328</td>
<td>36.85%</td>
</tr>
<tr>
<td>September 2009</td>
<td>163,992</td>
<td>11,153</td>
<td>15’17”</td>
<td>4,770</td>
<td>35.77%</td>
</tr>
<tr>
<td>October 2009</td>
<td>97,182</td>
<td>8,131</td>
<td>17’36”</td>
<td>3,816</td>
<td>32.66%</td>
</tr>
<tr>
<td>November 2009</td>
<td>92,945</td>
<td>7,676</td>
<td>16’32”</td>
<td>4,277</td>
<td>27.01%</td>
</tr>
<tr>
<td>December 2009</td>
<td>97,042</td>
<td>9,617</td>
<td>11’13”</td>
<td>5,243</td>
<td>28.92%</td>
</tr>
<tr>
<td>January 2010</td>
<td>102,441</td>
<td>10,521</td>
<td>13’50”</td>
<td>6,757</td>
<td>24.70%</td>
</tr>
<tr>
<td>February 2010</td>
<td>732,67</td>
<td>9,121</td>
<td>20’16”</td>
<td>5,654</td>
<td>20.17%</td>
</tr>
<tr>
<td>March 2010</td>
<td>173,569</td>
<td>12,056</td>
<td>15’31”</td>
<td>9,823</td>
<td>21.68%</td>
</tr>
<tr>
<td>April 2010</td>
<td>111,530</td>
<td>11,392</td>
<td>16’37”</td>
<td>5,104</td>
<td>23.75%</td>
</tr>
<tr>
<td>Average</td>
<td>267,325</td>
<td>8,917</td>
<td>16’17”</td>
<td>5,479</td>
<td>30%</td>
</tr>
</tbody>
</table>

As can be seen from Table 1, initial page views totaled 25,761 in May and reached 1,438,326 in August. Total visits were 3.2 million for the year. The number of unique visitors in May was 1,953, which increased to 7,314 in June after only one month. Although steady growth in unique visitors was observed in the subsequent months, the figure remained close to 9,000 people per month. In addition, there was an average of 5,479 returning visitors each month, accounting for 61% of monthly visits. This shows that frequent users are visiting the Platform regularly. Results also showed that users who access 20 pages or more per visit accounted for 30% of total visits.

The above statistics illustrate that the Platform is playing a role in the provision of government information services. In addition, we found that people visiting the site were from many regions around the country, and also from the United States, South Korea, Japan, Australia, and more than 30 other countries. Such wide use demonstrates the publicity and influence of the Platform.

As this information portal has only just started, however, it has not yet reached its full potential. Construction is still continuing and there are much work not started yet, especially in regards to external cooperation and publicity.
5 Existing problems

5.1 Lack of legal protection

There are no mandatory requirements on government agencies’ responsibilities and obligations for the public information listed in the Regulations. Therefore, libraries can not compel the government agencies to assist in the collection or provision of public information. At present, libraries can only make use of crawler tools for information collection. That significantly impedes the acquisition and release of information. As such, whether libraries can become qualified public service providers depends on the effective coordination mechanism between libraries and governments, a mechanism that relies on legal protection.

5.2 Lack of large-scale construction

Currently, only NLC collects government public information, which is inadequate for the total nationwide amalgamation of public information and services. Furthermore, it is very difficult for any single library to organize, manage and maintain public information from all levels of government agencies effectively. As such, the integration and provision of government information needs the collaboration of numerous different-sized libraries.

5.3 Technological difficulties

5.3.1 Information acquisition

Government public information was defined firstly as the main target resource during the preparation period of the project. However, content actually varies depending on specific government institutions, such as different public information of law and regulations, same category names but with different contents and improper or non-working web links. Those can cause difficulties in regards to information acquisition, and it requires a significant amount of time to clear data.

5.3.2 Information classification

Platform content is classified into twenty-two categories and only one hierarchy. This is unable to meet the needs of current users, and it is very difficult to find information by browsing for those whose retrieval skills are poor. We therefore need to optimize the classification scheme, allowing for additional sub-categories and greater classification dimensions, such as type of information. Although we made a preliminary attempt to do this, achieving the desired status through automatic classification was difficult due to the diversity and abundance of the information.
5.3.3 Data accuracy

Data accuracy is a fairly serious problem for the information acquisition via the Internet due to the dynamic nature of institutional information. For example, when government personnel changed the original website or removed a specific webpage, we were no longer able to access that information. When the crawler programs attempted to collect the webpage again, it was very difficult to accurately update the originally collected and published webpage.

5.3.4 Information preservation

This Platform uses two types of information preservation. One is plain text and the other is original webpage. The former is to only provide information retrieval and browsing of the contents and the latter is to preserve the whole original pages of the site including the style, the setup, and format of the page, etc. For certain special format pages, however, original preservation cannot be guaranteed.

5.4 Standards

Standards should cover whole the procedure including but not limited to information acquisition, organization, classification, preservation, and publication. As the Platform is still in its infancy, however, only metadata standards and classification schema have been applied.

6 Sustainable future development

After the promulgation of the Regulation of the People’s Republic of China on the Disclosure of Government Information in 2007, NLC planned and developed the three-year service platform project called Chinese Government Public Information Online. It was initially launched in early 2008, with the test version completed in December. On April 30, 2009, the platform was officially opened one year after the regulations were implemented. In addition, this project was unanimously approved by the Ministry of Culture office, public libraries, and experts of e-government during the Chinese Government Public Information Integration and Development Forum[2]. As seen from the statistical data presented in this paper, the service provided by the Platform is becoming increasingly important.

Given the wide variety of data sources, data complexity, and volume of government information, however, we are increasingly aware that it is impossible for any single library to organize and integrate all the available information in an effective scientific way. To address this issue, the Deputy Director of the Library, Chen Li, suggested to establish the National Public Library Union for Integration and Services of Government Information at the 10th Annual Conference of the Library Society of
China held in Nanning, Guangxi province, on November 3, 2009[3]. During the meeting, more than 300 representatives from all levels of public libraries and other institutions took part in the discussion, and 96 representatives from 82 public libraries signed the joint initiative. This laid a good foundation for the NLC to establish the government’s public information integration service union in China[4].

The overall objective of the union is to combine the public libraries of each province, city, district, and county in China and complete the integration and service of government public information with layered construction and resources sharing. The role of NLC is to integrate central government public information and for the libraries of each province, city, and county to integrate the government information of their own administrative region. From 2010 to 2011, the main members of the union will be the branch libraries of the National Digital Library. In February 2010, after research into the government information service needs of each branch library, NLC suggested that the existing Platform be extended and enhanced to be a united central platform, such that each union member library may use it to collect, integrate, and publish government information from their respective regions. Meanwhile, each union member can also establish their own site and issue their resources as a sub-station of the union. This will help accomplish the goal of layered construction and resource sharing and development of a personalized and unified display so as to provide a comprehensive government information service to the public. In April 2010, NLC conducted further research and analysis of the union’s Platform needs and optimized its system requirements. This proposed united central platform is now in the tender stage.

The launch of the united central platform is expected to occur at the end of this year. A detailed union work plan will be launched soon and other work to be carried out includes the following[5]:

- **Union construction** Complete the formulation of union policies, regulations, and service declarations; establish a special working group, and set up a union website and forum; establish the government public information service union, which covers most of the province libraries, as well as the city, district, and county public libraries in major provinces and cities.
- **Platform construction** Complete the construction and optimization of the union central platform, ask each union member library to collect and integrate government information from their respective regions, and issue them according to their own settings; ensure all resources on the platform achieve not only unified display and search, but also personalized display and search of different regions.
• Resource construction Collect resources hierarchically, build different databases through automatic metadata extracting and classification; analyze user needs to build special databases in fields that users are more interested in; effectively manage and preserve all resources in the long-term.

• Resource services Ensure the public can browse and search government information through this Platform; ensure the Platform provides basic resources of government information to each library that form complete service with other types of library resources and other service spaces.

• Training Provide government public information integration and service training to librarians of all levels of public libraries, volunteers, and union member libraries, and improve the government information service awareness and ability of librarians; carry out lectures on special topics, training, or other forms of education activities to enhance readers’ information literacy and the ability to make use of government information; carry out an investigation on readers’ information behavior to be the scientific basis for government information openness and service work.

• Government cooperation Strengthen the communication and cooperation with government departments in terms of information access, and raise the awareness of decision makers to public libraries; promote the construction of an open government information system, as well as the social and professional management of government information resources through close cooperation with government departments; increase the attention, reputation, and dependence of government on public libraries’ government information service.

• Public promotion Strengthen cooperation with organizations like news media, and open up new ways of promotion; formulate specifications and guidelines related to the organization of government information resources, promote demonstration projects of government information service in public libraries, strengthen the construction of government information integration service platforms and networks to encourage public libraries to play a full role in government information service.

• Academic communication Carry out academic exchange of government public information integration service, collaboration and resources sharing through seminars, academic forums and other forms, promote the communication and cooperation of government information services in public libraries of China, achieve certain consensus to improve the impact of public libraries’ government information service.

NLC strives to complete the collection of government public information of all provinces, cities, key districts, and counties in China during five years. By summing
up practical experience, follow-up development plans can be improved and implemented to realize the sustainable development of government information integration and service.

7 Conclusion

The construction of Chinese Government Public Information Online is an exploration and attempt of NLC to integrate and provide services on government public information according to the Regulation of the People’s Republic of China on the Disclosure of Government Information. This is a new information portal for government agencies, public libraries, and even the general public. There are still many issues that require further investigation and practice in terms of how to provide the public, government, and libraries with a more valuable government information service based on this Platform. Therefore, NLC will continue its efforts to not only encourage all levels of public libraries to form a National Public Library Union for Integration and Services of Government Information, but to also strengthen relationships with government departments and enhance communication with experts and scholars. It is important that all interested parties unite to provide an effective and integrated government information public service.

References


(Copy editor: Ms. Ning LI; Language revision: Chirstine WATTS)