Research and Application of Integration Solution for Enterprise-Level Heterogeneous Document Management Systems

Shujun Huang¹ and Xiangsheng Meng²

¹ Strategic Research Center of Oil and Gas Resources, Ministry of Natural Resources, Beijing, China
² CNPC Beijing Richfit Information Technology Co., LTD., Beijing, China
Email: huangsj@sinooilgas.org.cn; mengxiangsheng01@cnpc.com.cn

Abstract. This paper proposes an enterprise-level document management system integration solution, which can effectively solve the problems of the coexistence of multiple document management systems in enterprises and public institutions, the repeated collection of document data and the inefficient access application. The solution also can effectively support the data security and confidentiality of enterprises and public institutions. The core idea of the integration solution is to realize the application integration between the document integration application system and the document information management systems by taking the document classification collection management regulations as the management basis, taking the document information integration application system as the technical carrier, and adopting the SOA technology. The integration solution has been proved to be effective in practice, with significant application value and promotion value.

Keywords. Document management; system integration; SOA; service bus.

1. Introduction
In recent years, Strategic Research Center of Oil And Gas Resources, Ministry of Natural Resources (hereinafter referred to as the oil and gas center) has built a number of management information systems according to the needs of business development, which are oriented to mining administration support business, oil and gas resources strategic research work and internal general management. Among them, there are a number of systems related to document management, such as OA (office automation) system, scientific research management system, oil and gas resources information product release system, oil and gas geological data system, archives management system and internal and external websites. Due to the lack of a unified document management mechanism, there is a problem that multiple systems repeatedly collects and releases same documents. For example, the scientific research management system has collected the documents of scientific research achievements, but the oil and gas geological data system repeatedly collects and releases the documents. Documents managed by OA system are collected and published repeatedly on the Intranet website. The information products released by the information product release system are published repeatedly on the external website.

In addition, according to the regulations of security and confidentiality, different types of documents involve different management requirements and need to achieve hierarchical access. In different systems, it is necessary to carry out targeted access control according to the user's identity. All systems about document management need the document classification authorization and access control mechanism.
With the change of the user’s identity, all the relevant systems shall implement the authority change synchronously.

In addition to the above complexity of management, the separate operation of multiple systems also has a negative impact on the user experience. Users are not sure which system the documents they need in, and they cannot determine which one is the final version. They need to log into different systems many times to find the documents they need.

The problem is actually a relatively common problem. Many enterprises and institutions are facing the same kind of problem. In order to completely solve the above document management and application problems, it is much necessary to find a solution to eliminate the repeated collection of documents, unified query entrance, unified authority management.

2. Integration Solution Design

Through analysis and research, the project team successfully explored an effective solution. The core idea of the solution is shown in figure 1, including the following aspects:

(1) Improve the document collection and management regulations of the oil and gas center, establish a unified document classification model, identify the collection sources and positions of various documents, identify the specific document categories from various document management related systems, and control the repeated collection of documents by the management regulations and standards.

(2) Complete the collection and management of documents through the original document management systems. According to the management regulations, the document collection responsibility shall collect and store the documents in the corresponding category by the specified system as required.

(3) Establish an integrated document application system for the oil and gas center, provide a unified access portal for all users to realize unified access authorization and access control for all users of the oil and gas center. Improve user experience and ensure the compliance and convenience of document access.

(4) Adopt the service-oriented architecture to realize the integrated call of the integrated document application system and the original various document management systems. The SOA technology is used to eliminate the complexity of accessing multiple systems for users.

![Figure 1. Integration solution for enterprise-level documents management systems.](image)

The core functions of the integrated document application system are shown in figure 2. The system needs to realize four functions as follows.

System management: maintenance of user information, maintenance of document classification model, hierarchical access authorization for all users of different document categories and access control.

Document access: provide all users with a unified portal to document query, provide online browsing and package download function of all kinds of documents in the controlled state.

Statistical analysis: provide statistical classification for all documents of the oil and gas center, provide statistical classification of different document management systems. Provide the quantity statistics according to the collection responsibility. Specify the contribution of the different collection responsibility. Provide the comprehensive statistics of all kinds of document access and download to guide subsequent document acquisition optimization.

System configuration: Specify the document management system for integration access and search the time parameters of the document updates for each system.
3. Document Classification and Collection

Through in-depth analysis of all kinds of documents and management status, the project team lays down the document management regulations for the oil and gas center. The core contents of the regulation include the classification model of the oil and gas center document database, the document source, the collection responsibility and collection requirement and all kinds of document management system. Figure 3 shows the top-level classification of the oil and gas center document database.

**Figure 2.** Functional framework of integration document application system.

**Figure 3.** Document classification model.
4. Technical Implementation

4.1. SOA Technology

SOA (Software Oriented Architecture) is an architectural model that has evolved from component-based architecture, object-oriented architecture, and distributed systems [1-2]. SOA enables the integration of application components running on different platform systems and implemented by different technologies and programs [3-7]. It can achieve combinations of these loosely coupled application components. SOA links different functional modules in application program through interface to form serviced component [8-9]. The interfaces adopt the neutral way, and it is independent of the underlying hardware platform, operating system and programming language. It makes the services built in different system can use unified and general way to interact. Its essence is a kind of software component model with the guide of services.

This loose coupling nature of SOA provides the flexibility to build applications and business processes, either by using existing application systems as services, or by modularizing existing services or adding new services to meet new business requirements [10-11].

There are three main roles in the SOA:

1. Service requester: Use the services, through the service registry center to query the required service, and then according to its feedback information to bind to the service provider, then call the service to perform the service function. It is not necessarily a person, can also be an application program or software module.

2. Service registry center: Storage service, provide registration and query search function. It is an entity, not only allows service requesters to query searching operations through the network to get the address directory of the service, but also accept service providers to register and store services.

3. Service provider: Create the services and conduct strict testing. After the test, it passed the service to KI information published to the service registry center.

4.2. System Integration Implementation Based on SOA

The integrated document application system of the oil and gas center applies for the SOA architecture. The specific implementation adopts J2EE technical system, the Oracle database. The foreground page display adopts the combination of JSP and HTML pages, the background web container adopts Tomcat, and the enterprise service bus adopts Apache ServiceMix.

Figure 4 presents the technical implementation idea by an example of document access function implementation.

4.2.1. Develop Interface Services for All Document Management Systems That Need to Be Integrated.

The interface services for each document management system follow a uniform interface definition and mainly provide two functions:

1. The document integration application system obtains the document index information of the specified document management system. The document integration application system establishes the global document index database based on the document index information of each document management system.

2. The document integration application system obtains specific document files in the specified document management system.

4.2.2. Establish Global Document Index Database of the Oil and Gas Center. The document index update service automatically queries the interface service of each document management system according to the interval period specified by the configuration parameter. If the system document update list returned by the interface service is not empty, the global document index database will be updated based on the document updated index data.
4.2.3. Support All Users to Access All Kinds of Documents Uniformly. All users can search the documents based on the unified portal of the document integration application system. The document searching service is able to access the global document index database, obtain the relevant document list according to the query conditions entered by the user, and request the actual document file according to the user’s further view or download operation.

The document request service takes the index information of the document requested by the user as the service call parameter to the proxy service, which then passes the service request to the specific document interface service, which extracts the actual document file and returns it to the document request service according to the document index.

The document request service then passes the actual document file to the document request module or document presentation module to support the user’s document browsing or document download requests.

![Figure 4](image_url)

Figure 4. The technical implementation scheme of document integration application.

5. Application Effect Analysis

Through the development of the document integration application system for the oil and gas center, achieve the goal of enterprise-level document system integration. Not only achieve the inheritance and preservation of the original document management systems, but also provide a unified portal for the users to obtain documents efficiently and safely. It significantly improves the user experience. In addition, the solution realizes the unified authorization management of all kinds of documents, which guarantees the controlled access of all kinds of documents and effectively supports the data security and confidentiality work.

6. Conclusion

This paper proposes integration solution for an enterprise-level heterogeneous document management systems. It can effectively solve the problem of document repeated collection and inefficient access, and can effectively support the data security and confidentiality of enterprises and institutions. The integration solution has been tested by practice. The application effect is remarkable and the application and promotion value are significant.

References

[1] Xu H 2016 Research on the integration way of service oriented architecture Internet Technology 7.

[2] Xin C and Liang L 2016 Research on system integration based on service oriented architecture of J2EE platform Electronic Technology and Software Engineering 9 200-201.
[3] Zhou Z and Qu R 2016 Analysis and practice of-enterprise information system application integration technology China Management Informationization 3 75-77.

[4] Zhu M and Li Y 2017 Heterogeneous fusion mechanism with soa collaboration integration based on big data service deep requirements Computer Measurement & Control 25 (7) 165-169.

[5] Lin Y-J and Huangfu H-C 2017 Research on the integration scheme of the document management system based on SOA Electronic Design Engineering (18) 64-68.

[6] Wang P 2016 Research on digital campus resources integration based on SOA Information Technology (1) 159-161.

[7] Chen Y and Wang H 2016 Research on the mode of digital archives resource integration based on the SOA Archives & Construction (3).

[8] Jeong Y S and Shin S S 2016 An efficient authentication scheme to protect user privacy in seamless big data service Wireless Personal Communications 86 (1) 7-19.

[9] He Y 2019 Research on application system integration model based on SOA Information System Engineering 11 105-108.

[10] Fan X, Ju H, Wang M and Liu L 2019 Design and implementation of online service system for geological data based on SOA Architecture Environment & Engineering 2 280-284.

[11] Song Y 2019 Research and application of enterprise information system integration based on SOA architecture Electronic Technology & Software Engineering 3 256-258.