

RESEARCH AND IMPLEMENTATION OF CATALOG SYSTEM ON THE CONSTRUCTION OF THE GOVERNMENT INFORMATION CATALOG ARCHITECTURE

Wang Hu ^{a,b}, Wu Yiding ^b, Yang Xu ^b

^a Peking University, 100871

^b Datang Software Technologies Co.,LTD,100083

KEY WORDS: Catalog System, Metadata, Information Resource Sharing, e-Government

ABSTRACT:

The Government Information Catalog Architecture is one of the e-Government infrastructure, which is also plays a key role in developing and utilizing Government Information Resources. This paper focuses on the construction of the Catalog System from a technical perspective, introduces the requirements and functions of the system, and related technical details, which based on the author long-term research and practice experiences in related field. Some Conclusions are given at the last.

1. INTRODUCE

In recent years, the Chinese government has continued to develop construction of E-government, especially in the field of the Development and utilization of information resources. A series of documents have published to guide organization to create, manage and use Government Information Catalog, such as The Guidelines on Construction of e-Government in China by National Informatization Leading Group (General Office, CCCPC, no.17, 2002), The Opinions for Improving the Development and Utilization of Government Information (General Office, CCCPC, no.34, 2004), National e-Government General Architecture (State Council Informatization Office(SCITO, no2, 2006), The Strategy for National Informatization from 2006 to 2020 (General Office,CCCPC,no.11,2006). All above show the construction of the Government Information Catalog is one of the most important issue for the success of the construction of e-Government.

Government Information Catalog (GIC) is build upon the National e-Government Network, which covers the state, province, city and town, forms a multi-level catalog layout.. GIC uses metadata to describe and unify the sharing of information, it organizes and manages the information resource which can be looked up and located in a unified logic-view. GIC also supports plenty ways for information resource sharing between different offices and districts and councils of the state government, which provides GIC-user information resources

effectively and efficiently.

The construction of the GIC need cost tremendous efforts in many aspects such as policy-making, standard-defining and system-building. The system-building contains the building of the Cataloguing System and the building of the Catalog System, and also includes the building of the related Database System. This paper focuses on the construction of the Catalog System from a technical perspective, introduces the requirements and functions of the system, and related technical details, which based on the author long-term research and practice experiences in related field. Some Conclusions are given at the last.

2. OBJECTIVE AND REQUIREMENT

The requirements of the Catalog System are closely related with the Cataloguing System. Both of them are the key members of the construction of the Information Catalog Architecture. There are three phases of the Catalog Architecture process, which includes Cataloguing, Catalog management and Catalog inquiry. The main business entities for the three phases are Metadata cataloguing unit, the catalog center and the catalog user. The cataloguing system is the information collector, which provides standardized metadata to the Catalog System. The Catalog System provides catalog inquiry service to the Catalog user.

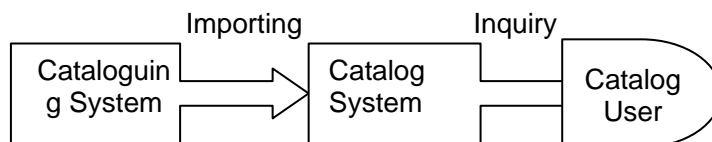


Figure.1: the main business process of the Catalog Architecture

The business process in the figure.1 shows that the Catalog System receives the metadata of government information which provides by Cataloguing system. The Catalog System Audits and verifies the metadata information and import it into

database. The catalog System publishes the information catalog, and then the catalog user can query the catalog system to locate the desired information resource.

3. FUNCTIONAL ARCHITECTURE

The thorough system functional architecture view shows as the following figure:

Depends on the system requirement, the Catalog System contains three sub-system. There are the Metadata Registry/Publish Sub-system, the Management and Maintenance Sub-system, and the Catalog Service Sub-system.

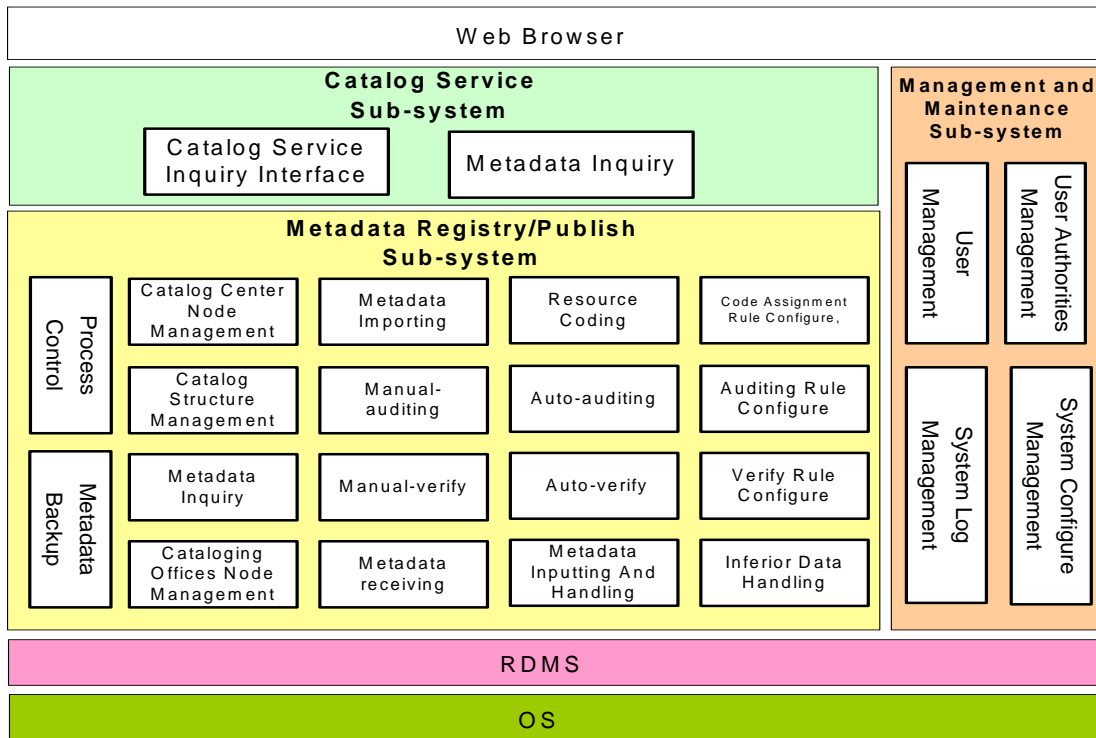


Figure.2 System functional Architecture

■ **Metadata Registry/Publish Sub-system :** Process Control, Catalog center Node Management, Metadata Importing, Resource Coding, Code Assignment Rule Configure, Catalog Structure Management, Manual auditing, Auto-auditing, Auditing Rule Configure, Metadata Backup, Metadata Inquiry, Manual-verify, Auto-verify, Verify Rule configure, Cataloguing Offices Node Management, Metadata receiving, Metadata Inputting and Handling, Inferior Data Handing.

■ **Management and Maintenance Sub-system:** User Management, User Authorities Management, System Log Management, System Configure Management

■ **Catalog Service Sub-system:** Catalog Inquiry Service and Service Interface

4. TECHNICAL ARCHITECTURE

4.1 Lightweight Application Development Framework

The system technical Architecture is based on the industry standard J2EE specification. It is built upon industry best practices and constructed as an Distributed Multi-tiered Web application with the important supporting technology such as Distributed Component and Web Service and XML related tech suit.

System Development framework is using Appfuse framework. Appfuse is an open source J2EE application development framework that uses lightweight container (The Spring Container) which built on the J2EE platform to help you develop J2EE Web applications effectively and efficiently. The following figure shows the Catalog System Technical Architecture with Appfuse framework.

The Technical Architecture consists of the three tiers shown in Figure.3: the Presentation Tier, Business Logic Tier and the Persistence Tier.

1) **The Presentation Tier:** the presentation tier handles user requests and generates HTML or XML, and other Data content for responses. It uses the classic J2EE presentation tier technology such as JSP and Servlet technology and J2EE presentation-tier Design Pattern as Front-Controller.

2) **The Business Logic Tier:** the business logic tier handles domain-related business logic and controls transactions and interfaces with other tiers. it manage the objects' decencies on the business level.

3) **The Persistence Tier:** the persistence tier is used by the business tier to access external Catalog database. It handles the persistence of the business objects.

With the help of Appfuse framework, System confines the special functional module into a pro-defined level to effectively remove the functional redundancy.

Catalog System left all the business components to the Spring Container, which manage the whole life cycle of all the business

component objects. It's shown in the Figure.4.

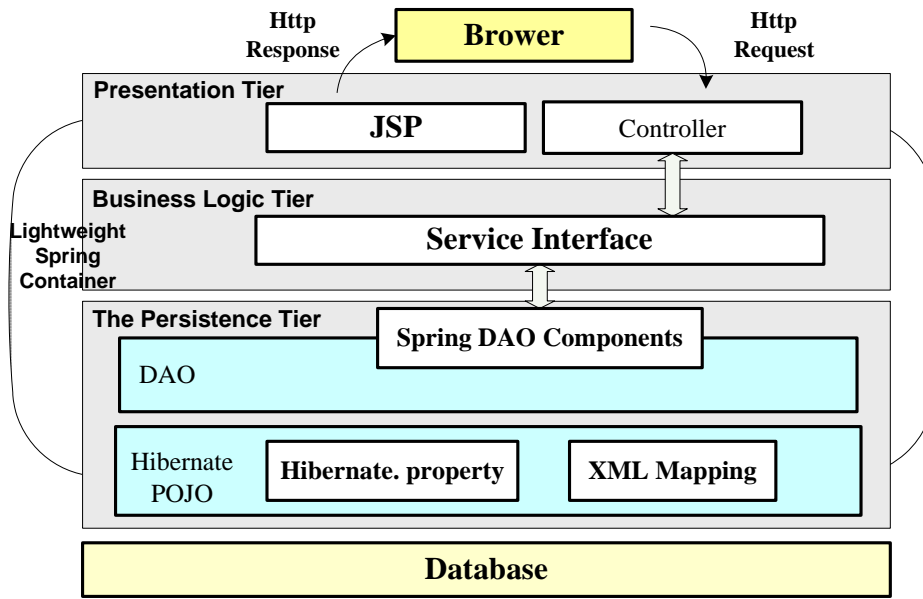


Figure.3 System Technical Architecture with Appfuse framework

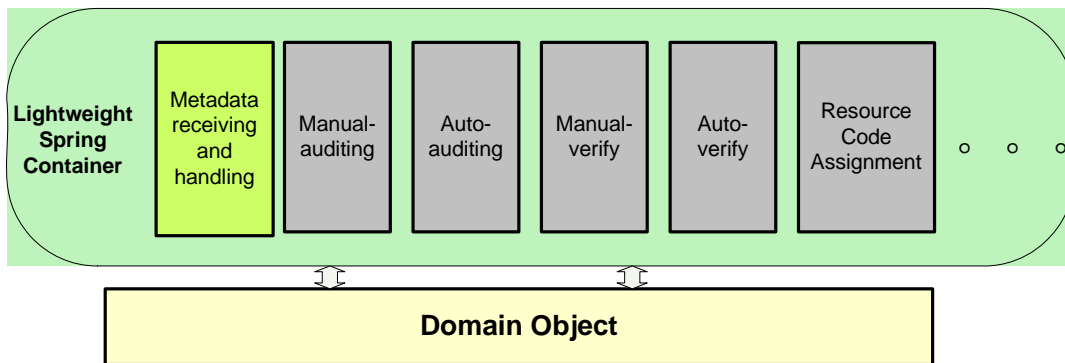


Figure.4 Business Components Managed by Container

As shown in figure.4, Business Components such as Metadata receiving and handling, Manual-auditing are managed by spring container. *Metadata receiving and handling* provide service interface to cataloguing system using Web Service Technology. With the help of the spring Container, the module auto-wires into system without side-effects to the other modules inside the spring container.

4.2 Catalog Database Design for Multi-standards

The metadata and its related standards and specifications are the fundamental elements on the constructions of the Catalog system. Metadata describes the Information in a specific way and it can be recognized from the information resource to form a characteristic element collection. Using metadata is a simple and reasonable way to manage, search and acquire information, people can recognized all kinds of details about the information resources, such as its format, quality, and handling and acquire method and so on. Metadata standard can be used in multi areas include Resources Sharing, Data Publishing, Data Set Cataloguing, Data Exchanging, Data Inquiring and so on, Metadata standard is the standard structure of metadata data set for building database , assembling and publishing metadata

There exists different Metadata Standard in different field, such as The Metadata Standard of Geographical Information, The Core Metadata Standard of National Government Information. The Catalog Database structure need be designed to meet a wider range requirement to support multi-standards.

4.2.1 The fundamental for catalog database design:

Database model will target on supporting functional requirement for Metadata store and Metadata Inquiry, such as:

- (1) Must exist a unique Identifier to identify every single metadata record stored in the Metadata database and only unique identifiers list should be returned when execute an metadata inquiry.
- (2) Should exist a fast and direct way to return an whole metadata data record (XML file) by an specific unique identifier.
- (3) Can import metadata data resources (in a valid XML format) into metadata database through a interface

4.2.2 Support for a variety of metadata specifications: The Catalog database is designed for supporting metadata standards and specifications over many domains. There are two data

entity: metadata database description table and metadata standard description table. The first table stores the information about metadata database, includes: metadata database identifier, name, description, metadata standard, metadata table name and so on. The second table stores XML schema information about metadata standards and specifications includes: metadata standard identifier, name, version, description, standard content and so on.

The metadata standards stores in the metadata standard description table and every metadata standard stands for a record of the table. The metadata information for this kind of standard stores in a group of tables, which include Metadata content table, Metadata element table and so on.

When the System need to adapt to a new metadata standard, the design introduced above keep the things simple: just Adding a new record to metadata database description table and metadata standard description table and adding a group of data entities to

store this kind of metadata.

5. THE SYSTEM FUNCTIONS

5.1 The Metadata Registry/Publish Sub-system

The Metadata Registry/Publish Sub-system receives the Metadata coming from the Cataloguing System, audits and verifies the Metadata, and import the verified and valid Metadata into the Metadata database, and backup the importing Files which contain the Metadata. The following figure shows the workflow of the process.

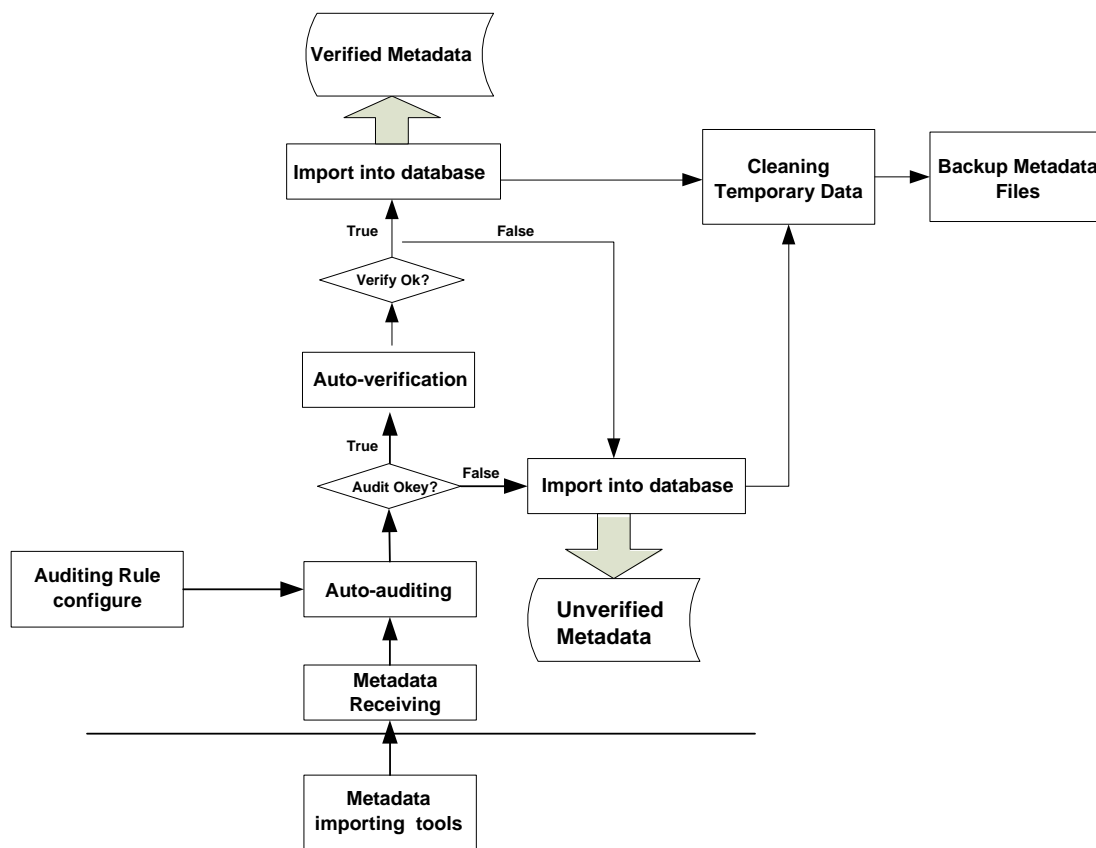


Figure.5 Metadata Registry/Publish Workflow

The Metadata Registry/Publish Sub-system manages the catalog center and cataloguing work-units as nodes. We can create a new node, and modify or update the details information of any specific node. Registry/Publish Sub-system contains auditing rules, verification rules and code assignment rules, which can be used to keep the metadata importing process automatic and being controlled. The Sub-system also support manual auditing and manual verification, user can lookup the metadata and decide if it can be polished.

The Catalog Structure Management can create and maintain catalog layout. It supports import and export the catalog

structure, and create and display the catalog layout.

5.2 Catalog Service Sub-system

The Catalog Service Sub-system acts as a role of inquiry services provider. The Metadata Inquiry Module receives the inquiry requests from the Catalog Inquiry Service Interface, then it search the metadata database and return the result set to the Inquiry requestor. The sub-system support a Variety of inquiry ways includes basic inquiry, advanced inquiry, catalog inquiry and so on. The inquiry output format conforms to standard XML format.

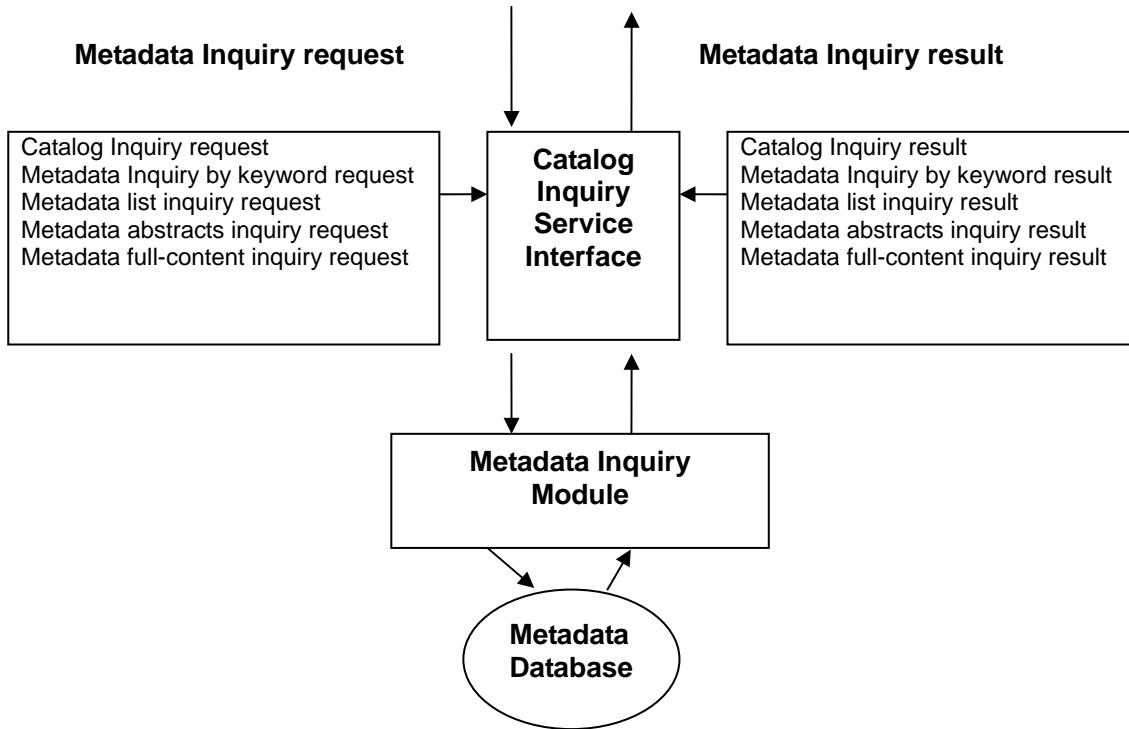


Figure.6 Catalog Service Sub-system inquiry workflow

In this sub-system, includes following functions:

1. Provides overall searches, which support users (or other systems) searching all the metadata databases in the catalog system via network.
2. Provides different access levels, which constrains user searching results by user accessing level.
3. Provides information abstracts, which useful for user to identify a specific metadata easily and quickly.
4. Access protocol bases on HTTP protocol, and can be extended to support SOAP easily.
- (3) The Metadata Auditor, which manages the content of the metadata in the metadata database and audit the received metadata from the metadata registry system.
- (4) The Metadata Register, which registers metadata to the metadata catalog services system.
- (5) The Statistics Viewer, which views the statistics during the system running, the statistics includes metadata amount, metadata distribution, access status, and system running status and so on.
- (6) The Common User, which search and browses the catalog and metadata of the system. The common user can also be used for anonymous access.

5.3 Management and Maintenance Sub-system

The Management and Maintenance Sub-system provides system management and maintenance functions for system logs, user information, user authentication and system configuration.

The sub-system defines role-based authentication, users are assigned one or more predefined roles. The predefined roles include the system administrator, the catalog administrator, the metadata auditor, the metadata register, the statistics viewer, and the common user.

- (1) The System Administrator, which manages and configures the system services and the metadata database, and maintain other roles and users of the system.
- (2) The Catalog Administrator, which defines, manages and maintains the system metadata catalog structure and generation rules.

The System Configure Management module is in charge of managing and maintaining system perimeters required for system running. The System Logs Management module is in charge of managing system operation logs, backing up the log files by time periods and log types, and viewing of the historical logs.

5. SYSTEM DEPLOYMENT VIEW

The Construction of the Government Information Catalog Architecture is based on the National E-Government Network. The Catalog System Multi-levels layout covers States, province, city and town, which forms a physically dispersed and logical centralized Information Resources Sharing Model.

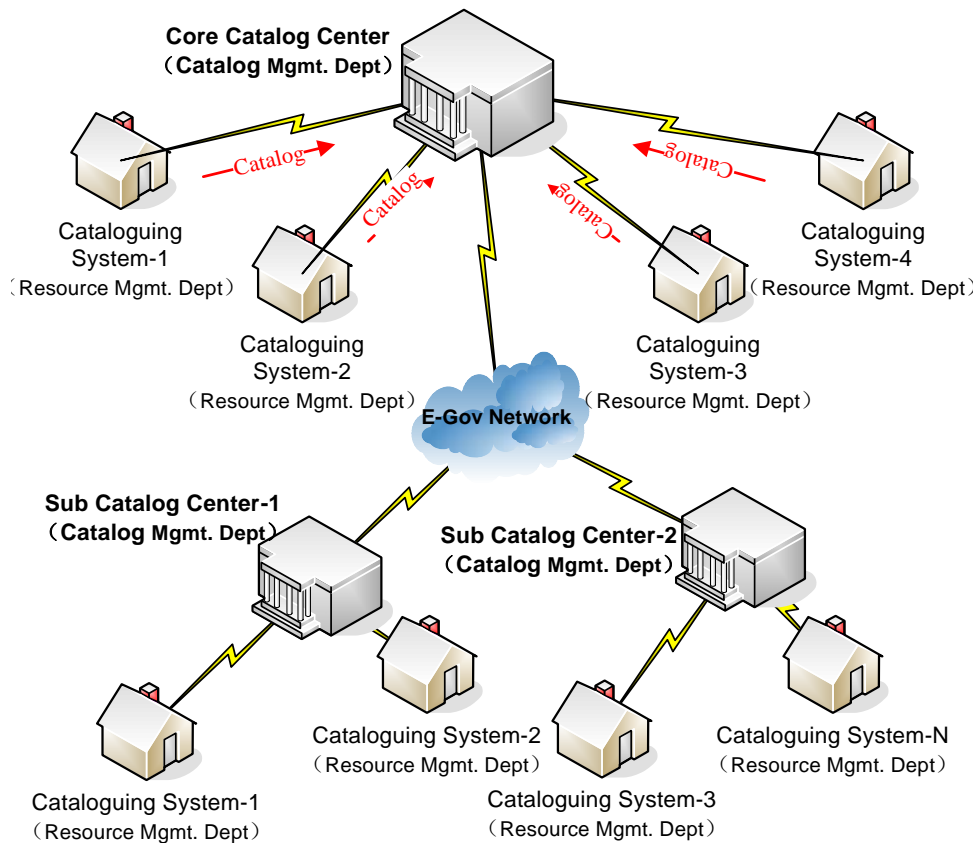


Figure.7 Deployment view of The Catalog System

In an actual system deployment view, the Catalog System contains two levels: the Core Catalog Center and the Sub-Catalog Center. The Catalog Center is the fundamental node of the architecture, which usually is Information Resources Management Center located in a city or a province. The Core/Sub Catalog Center contains the catalog system and related metadata database. The Core Catalog Center can access the Sub Catalog Center. The database in the Core Catalog Center not only can store the metadata coming from the cataloguing system for the same level (shown in figure.7) but also the metadata from the sub catalog center. The Cataloguing systems are deployed in the Resource Management Departments. The cataloguing work-units import local information resources using cataloguing systems, and register to the catalog systems.

6. CONCLUSION

The construction of the Government Information Catalog (GIC) is the very important aspect of the Government Information Resources Sharing in China. Recently, National Standards of Government Information Catalog will be released. The related pilot projects leading by State Council Informatization Office (SCITO) are on schedule in Beijing and in Inner Mongolia Autonomous Region. All these efforts will further Strengthen and improve the development of GIC. The Catalog Architecture is playing an increasingly important role in the Construction of GIC.

There are some factors worth serious considerations in constructing the Catalog Architecture, especially to the issue of management, standards, and technologies. Related research will

continuously.

REFERENCES

- Lai Zhibin, Wang Hu, Research of Application Model Based on Management of Metadata and Dataset, Progress in Geography, Vol.21, No.4 (July, 2002)
- Mu Yong, Wang Hu, Government Information Catalog System, Local Standard in Beijing, April, 2006
- Wang Hu, Li Qi, Building and Implementing Metadata Services Architecture of Digital City, Universitatis Pekinensis, Acta Scientiarum Naturalium, Vol.40, No.1 (Jan, 2004)
- Wang Hu, Research on Geo-Ontology Based Metadata Service Architecture and Technology, Doctoral Dissertation, Beijing: Peking University Library, 2003.5