

# A metadata based agricultural universal scientific and technical information fusion and service framework

Cui Yunpeng<sup>1</sup>, Liu Shihong<sup>1</sup>, Sun SuFen<sup>2</sup>, Zhang Junfeng<sup>2</sup>, Zheng Huaiguo<sup>2</sup>

<sup>1</sup> Key Laboratory of Digital Agricultural Early-warning Technology, Ministry of Agriculture, Beijing, The People's Republic of China 100081

<sup>2</sup> Agriculture Sci-Tec information institute of Beijing Academy of Agriculture and Forestry Science, Beijing, The People's Republic of China 100097

**Abstract:** The paper introduced a metadata based Agricultural scientific and technical Information fusion and Services framework. Through Agricultural scientific and technical Information dataset core metadata and Services core metadata, the distributed and platform-independent Information fusion can be implemented, based on the information fused resources in base layer of the framework, many applications can be developed, such as mobile communication based mobile Information service, voice text converter based voice information service, smart Q & A application etc., and the solution is an available and effective solution for the fusion and services of agricultural scientific and technical information resources, because the solution can integrate the data with different format from different data sources, so the solution can be used to construct the data layer of agricultural scientific and technical universal information services.

**Key words:** Information fusion, Metadata, Agricultural scientific and technical Information service

## 1. Introduction

The construction of Agricultural informatization in China has made rapid progress in recent years. There are now over 2,000 E-commerce web sites, more than 6,000 Agricultural web sites in China, and a large amount of application and digital products have emerged<sup>[1]</sup>, the agriculture productivity in China has been enhanced significantly. But on the other hand, some problems also emerged during the period. because there are too many information resources located in different data sources, and a lot of these information are duplicated, it's very difficult for users to find the right knowledge and information they really want, and users can't extract useful knowledge from just one or several information. The crux of all these problems is information fusion, based on information fusion, the information resources with different formats from different data sources can be integrated in one logical whole, the retrieval to this logical whole is just like retrieval from one dataset, the duplicated information entities can be removed automatically, based on the fusion of the information resources, many applications can be developed, such as mobile service, intelligent Q&A system, the cross- database retrieval application etc..

The most effective scheme for information fusion nowadays is metadata, it is the core of information fusion, and now there are already some successful case which use metadata based framework in information fusion, such as the scientific database service platform of CAS, the agricultural scientific data sharing platform of CAAS, etc., this paper will discuss a metadata based agricultural scientific and technical information fusion and service framework as well as the standards worked out for the framework.

## 2. Metadata and metadata standards

The definition of metadata is: metadata is the data about data<sup>[2]</sup>. In fact, metadata is a set of data that can describe and identify a specific information entity, and help users find and achieve related information resources object. There are two kinds of metadata standards, core metadata standard and expanded metadata standard, normally, expanded metadata standard was expanded from core metadata standard and used in some specific area.

The usage of metadata standards include:

1. Information management. Metadata can describe information entity, that means all the information entities can be “labeled” with metadata, through metadata, the management of information entities can be promoted.

2. Information discovery. With metadata, the retrieval of information entities is in fact the retrieval of metadata registration information, so the metadata registration information can be regarded as the sources of information discovery<sup>[3]</sup>.

3. Information acquisition. Normally, the position and the type Information is included in metadata, so users can acquire information entities very easy, thus users can utilize information more effectively.

4. The integration and sharing of information resources. All the information entities is registered with universal or compliant metadata standards, so the integration and sharing of information resources turns into reality<sup>[4]</sup>.

## 3. Metadata and information fusion

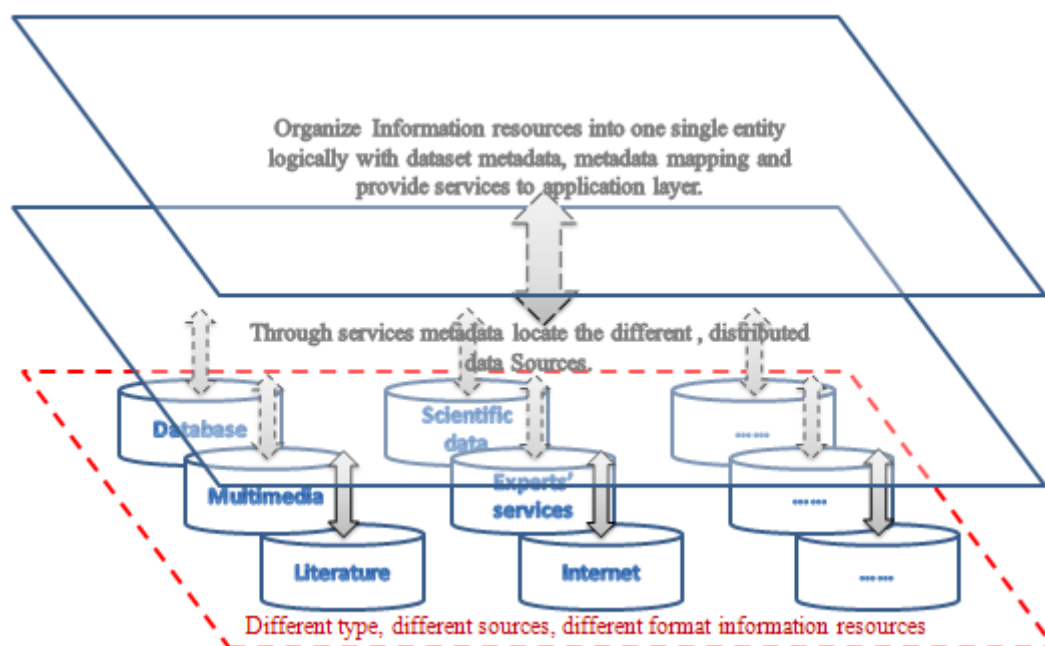


Figure 1 the framework of information fusion with metadata

Figure 1 shows the principle of information fusion framework with metadata. There are 3 layers in the framework. at the bottom of the framework is data sources layer, where contains different type, different format information resources from different sources distribute anywhere, such as databases, multimedia resources, literatures, scientific data, Internet resources etc.. The middle layer is services metadata fusion layer, in this layer, all the data sources at the bottom layer are identified by services metadata, this layer include a services metadata value database, where all the data sources information saved, users can locate a data sources through these services

metadata values and connect to a data source to get the information they need. The top layer of the framework is dataset metadata fusion layer, all the datasets locate in the data sources at the bottom layer will be identified by dataset layer, so there is a large dataset metadata value databases in this layer, users can retrieve the information entities in the datasets through retrieving the dataset metadata values.

If a user want to retrieve an information entity, he/she enter keywords into an application that provides the services above, the system will first retrieve the information in the dataset metadata values database, find which dataset include the information entities match the keywords the user entered, and then locate the data sources where the service metadata described, find the connection information of the data source, then connect to the data source, send query information, get the results, and feedback results to the user.

#### 4. The agriculture information resources dataset core metadata

The agriculture information resources dataset core metadata is a metadata standard which regards agriculture information resources as descriptive objects, it is expanded from the scientific database core metadata standard of Chinese academy of science. It defines a set of data modules and elements. The main body of the standard includes 6 required modules: dataset description information, data quality information, dataset distribution information, metadata reference information, services reference information, and structure description information, and 2 assistant modules: range information and contact information. The assistant modules can only be quote by the elements of the required modules, and cannot be used separately, See Figure 2.

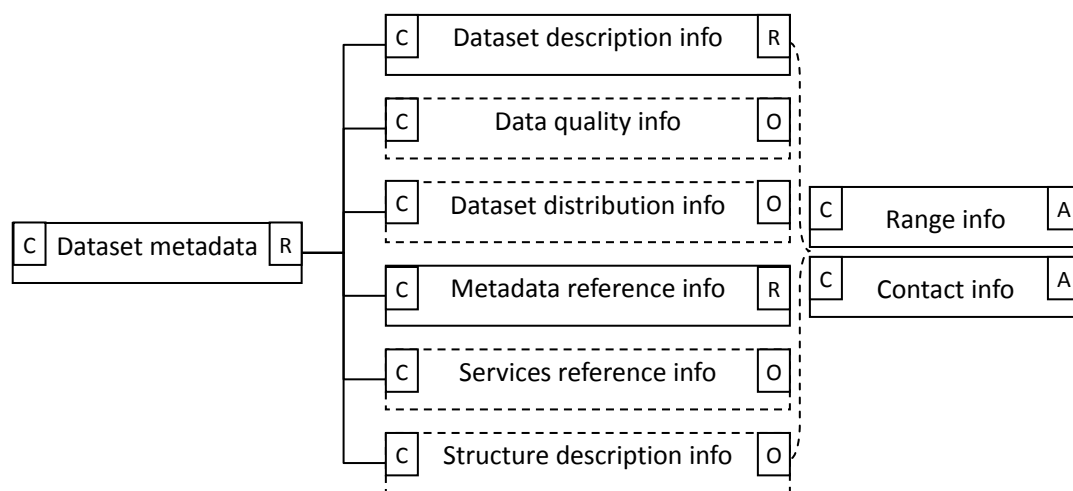


Figure 2 the structure of the agriculture information resources dataset core metadata (C composite module, R required module, O optional module, A assistant module, the same hereinafter)

All the elements in the standard has 9 attributes, as shown in Table 1.

Table 1 the attributes of the elements of agriculture information resources dataset core metadata

Name of attribute	Description
Chinese name	Chinese name of the element
English name	English name of the element
Identification	The unique identification of the element, string.
Definition	The specifications description of the meaning of the element.
Type	The type of the element, the available types include: composite(the element contains sub

	elements),integer, float, text, date, time, datetime etc.
Range	The allowed range of the value of the element
Optional	The element is required or optional
Maximun appearance	The maximum appearance of the element, such as 1 (only once) 、 N (unlimited times) etc.
Note	Supplementary specifications of the element

### 5. Agriculture information resources services metadata

The agriculture information resources services metadata defines specific services metadata specification. For a specific service, the metadata specification is relative fixed, so we can find a model to define any services. Figure 3 shows the universal model of agriculture information resources services metadata.

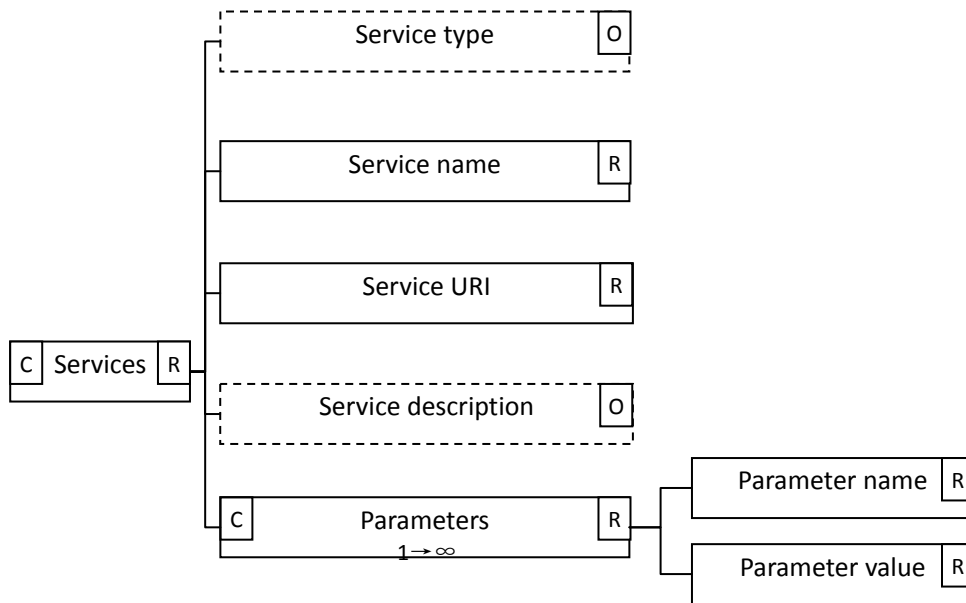


Figure 3 the universal agriculture information resources services metadata model

The universal model includes 5 elements: service type, service name, service URI, service description and parameters, the parameters can be one or more, each parameter has parameter name and parameter value.

For example, the dataset connection service metadata is like the following (see Figure 4 ).

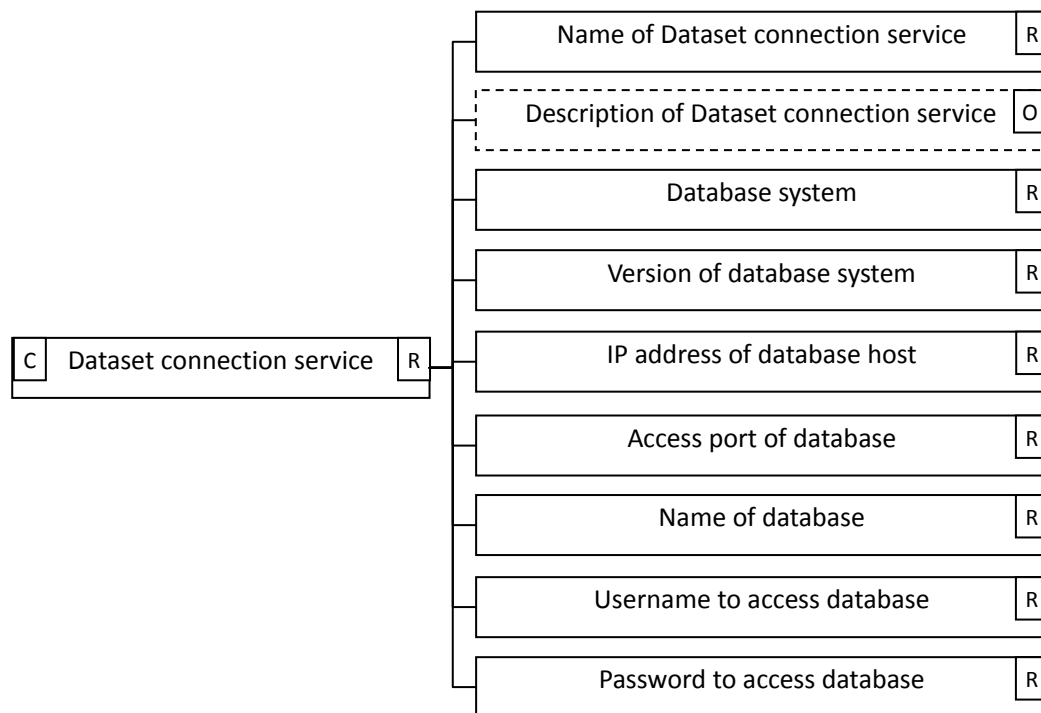


Figure 4 the example of dataset connection service metadata

From the example, we can see that all the information needed to connect to a database are contained in the metadata values. That is, an application can connect to the database automatically with these information, once an agriculture database is indexed with agriculture information resources dataset connection service metadata, the application can connect to the database and get results from database, the whole process will be finished automatically.

## 6. Conclusion

Information fusion is very important for agriculture scientific and technical information services, but it's very difficult to find an effective mechanism to implement real agriculture information fusion. Metadata provides available means to integrate different type, different format information resources from different sources, and all the information sources can be integrated into one logical entirety. Based on the integrated information resources, different application can be developed, such as mobile communication based mobile Information service, voice text converter based voice information service, smart Q & A application etc., so the universal information services can be implemented, thus, the quality of agriculture information services will be promoted greatly.

## ACKNOWLEDGEMENT

The research was supported by the special project from ministry of agriculture of the people's republic of China, named study of agriculture informatization standards system and special fund of basic commonweal research institute project of information institute of CAAS, and National 11<sup>th</sup> five-year technology based plan topic named study of Agricultural product quantity Safety Data obtained standards (2009BADA9B02).

## Reference

- [1] Zheng Huaiguo, Tan Cuiping. The Integration and sharing of Agricultural Information Resources in network environment. Journal of Anhui Agriculture Science.2008,36(13):5665-5668
- [2] Xiao Long, Chen Ling. Chinese Metadata Standard Framework and Its Applications, Journal of academic libraries. 2001,19(5): 29-35
- [3] Michael Jennings, David Marco. Universal Meta Data Models. Wiley Publishing, Inc.,2004
- [4] Qian Ping, Su Xiaolu,Cui Yunpeng. Study on agricultural scientific and technical information core metadata.Agriculture Network Information.2006,(2):18-21