

Case study: "The building blocks of a Forestry Spatial Information Catalog (FSIC)"

Joris Siermann - Systems Architect FSIC
Center for International Forestry Research

The Center for International Forestry Research (CIFOR) has developed a catalog service to search, find and retrieve forestry-related spatial information using ESRI tools (ArcCatalog, ArcSDE and ArcIMS). The Forest Spatial Information Catalog architecture includes components for metadata generation (based on the FGDC standard), OpenGIS compliant Web Mapping and client side metadata searching and retrieving. The entire content of the database has been made available to authorized users; the searching and viewing of the metadata is open to everyone. This paper describes the design and implementation of the various components and how this service was introduced in an international research organization.

The problem

The volume of forest-related information generated globally is enormous and the number of sources of information is equally overwhelming. Increasingly, such information is published using electronic means, but a substantial part of it remains inaccessible to users. The major problem facing information-seekers is the location of information sources corresponding to their needs.

Equally important, information-providers often do not easily find appropriate media for presenting their information, which then remains inaccessible to others. It is significant to solve the problem of "hidden" data; accessible data will stimulate the efficient use of it and therefore removes the redundancy of information and improves the quality of the work performed by CIFOR scientists.

CIFOR is in the early stages of addressing this issue. Previous work was done on defining an applicable metadata standard and setting up a pilot meta-database for spatial data. The successes gained by these efforts were a common CIFOR metadata description standard, knowledge about metadata practices and a start of describing some spatial data. The pilot used the CIFOR library software which was not fully catered to fulfill the needs of a geospatial information management system. Furthermore, the system was not updated and maintained so the limited number of users started to abandon the initiative. Now with new technologies and software that is acquired as a result of a memorandum of understanding between CIFOR and ESRI, new opportunities arise to re-address this issue.

In order to cope with the new tasks, the GIS Unit defined two main goals, (1) to maintain providing quality GIS support to the research divisions and (2) to improve spatial data accessibility by establishing a spatial data management system, based on international standards and best practices to support a wide range of forest related geo-spatial sources and to stimulate stakeholders to provide data.

In order to achieve these goals, we propose objectives in the field of (1) *Infrastructure*; which relates to the design of the spatial data management system, (2) *International Standards*; referring to compatibility with standards like the Federal Geographic Data Committee (FGDC) and OpenGIS as well as other catalog initiatives like the Global Forest Information Service (GFIS) and the Consultative Group on International Agricultural Research's Consortium for Spatial Information (CSI), and finally (3) *dissemination*, which refers to the important factor of user stimulation by means of raising awareness, training and knowledge sharing of the new system.

User needs

Making (geo-) information available to the whole organization usually takes more than just the implementation of the appropriate software. The organization itself needs to be adjusted to it. Therefore, the first step in designing a knowledge based information system is to get a thorough understanding of the users of the system. In the case of CIFOR, the user needs could be divided into two groups; the first related to the institute, the other related to the individual users.

Institutional requirements

All data produced by CIFOR is regarded as a global public good. CIFOR has therefore the institutional obligation to disseminate its knowledge. From CIFOR management point of view, a strategic instrument was needed when it comes to the dissemination and retrieval of their valuable (geo-) information. This resulted in certain quality assurance procedures, separation between the actual data and the information service (security) and guidelines with regards to intellectual property rights.

Individual user requirements

The individual users of the FSIC are the CIFOR scientists, as well as a diverse set of stakeholders at local, national and international level (government analysts, conservation and development NGOs, producer organizations, researchers) who need forest-related geospatial information for various purposes, such as land use planning, forest resource assessment, carbon trading and biodiversity conservation. Questionnaires and interviews were held with the aim to define a common goal and to find out what was wanted or needed. Results are summarized in Table1.

Table 1: Result of user needs survey sorted from highest priority to lowest (source M. van Heist 2002)

Issue raised	Proposed activity	Priority (Y/N)
What exists?	Create inventory	Y+
Coordination	Identify coordinator	Y+
Priority needs	Inventory and structured archiving	Y+
Access to information	Create metadata	Y+
Management	Collecting info, archiving, updating	Y+
Access to data (CD-ROM, printout)	Create metadata	Y
Acquisition and updating	Increase contacts with other organizations	Y
Current management	Determine responsibilities	Y
Quality assessment	Critical review	Y
Access to raw data		N

System design

The key aspect of the system design is to provide answers with regards to the data quality, data entry procedures, tools and procedures for system administrators as well as guidelines on how to publish metadata for online data-sharing.

Infrastructure Objectives

The overall infrastructure of FSIC consists of components related to the data and meta-data storage (database with ArcSDE), the GIS tools for data analysis and reporting (ArcGIS/View) and a web portal for online access to the meta-database (Figure 1).

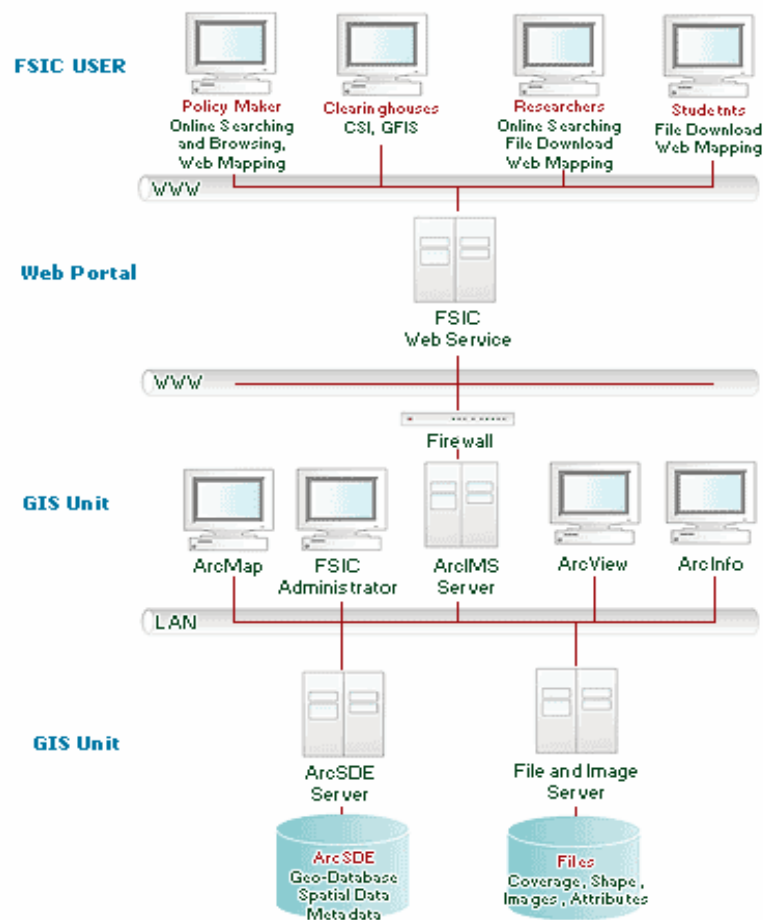


Figure 1: the FSIC system architecture including the web portal

The FSIC web service

The FSIC web service provides online access to the geo-spatial search tool with has direct access to the spatial meta-database. We have chosen to develop the service using Java technology and the Open source STRUTS web application framework created as part of the Jakarta project at Apache. The Java/STRUTS technology enforces the Model2 MVC design pattern, making it hard not to follow the correct software design. Applications developed this way take a little longer to create up front, but the maintenance is a comparative breeze.

Table 2: Technology used in FSIC

ESRI ArcIMS and ArcSDE
Java 2 Edition
Java Servlets, STRUTS, JSTL
Apache Tomcat Servlet Container
Apache 2.0 HTTP server
Microsoft SQL server
Microsoft Windows 2000 Server

Metadata

At the moment many changes are taking place in the field of meta-information, for instance the completion of the CEN standards and the activities around the ISO standards. These changes are crucial to the future direction of meta-information developments. In order to

increase the interoperability of (geo-) information and meta-information, an information management system should make use of the current standards in the field of meta-information. For the implementation of the catalog we decided to go for the FGDC metadata standard (Federal Geographic Data Committee) because it facilitated all the needs we had with regards to the metadata entry fields and it was already implemented by the ESRI product line.

Searching metadata - logical work flow

In essence, the processing requires users to interact with web pages and enter information after entry. The underlying work flow passes through all the layers of the architecture, interacting with all the components necessary and returns the result to the user in a resulting web page (Figure 2).

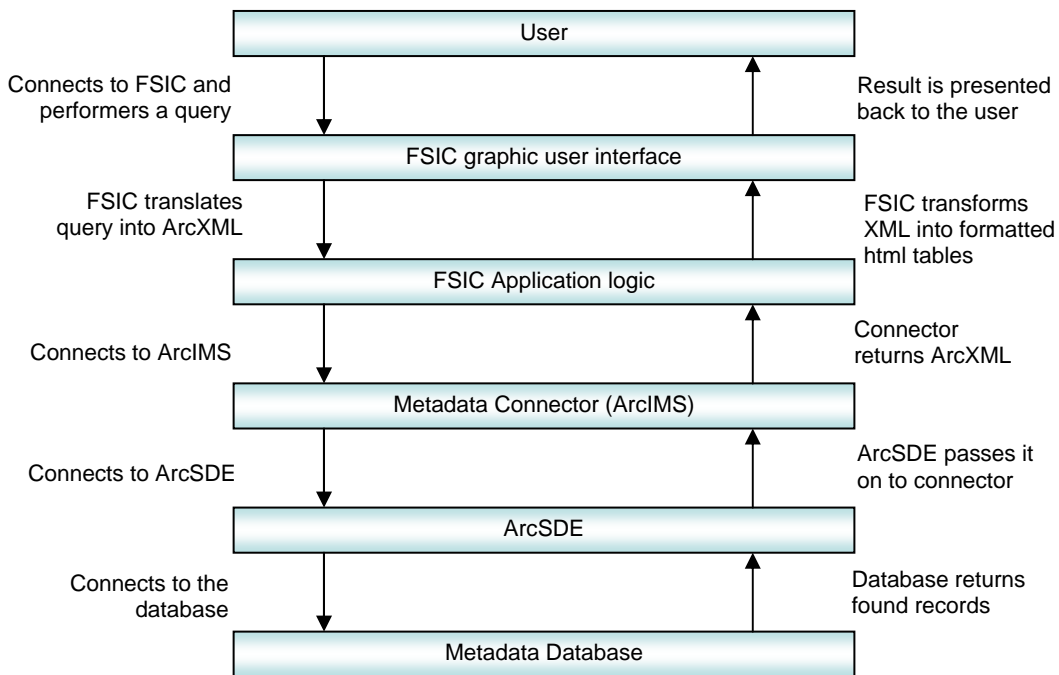


Figure 2: a typical FSIC user activity workflow

Publishing metadata - logical work flow

The generation of metadata starts with a collection of spatial datasets which is either a shared location on a file server, a spatial database or hardcopy maps. The GIS specialist uses ArcCatalog to select a data source, subsequently calls a metadata entry application and start to enter or update the metadata. When finished, the specialist connects to the ArcIMS metadata connector using a username and password and after a successful login procedure, the metadata can be published in one or more predefined metadata categories (Figure 3).

Data access policy

Basically, all the data maintained by the system is protected by the regulations of the CGIAR centers as mention on the document "Legal Issues in the Use of Geospatial Data and Tools for Agriculture and Natural Resource Management". Most of the CGIAR data is regarded as a global public good, which means that its information should be available for everyone. However, in some cases, access to sensitive or otherwise not public data is restricted to authorized users only.

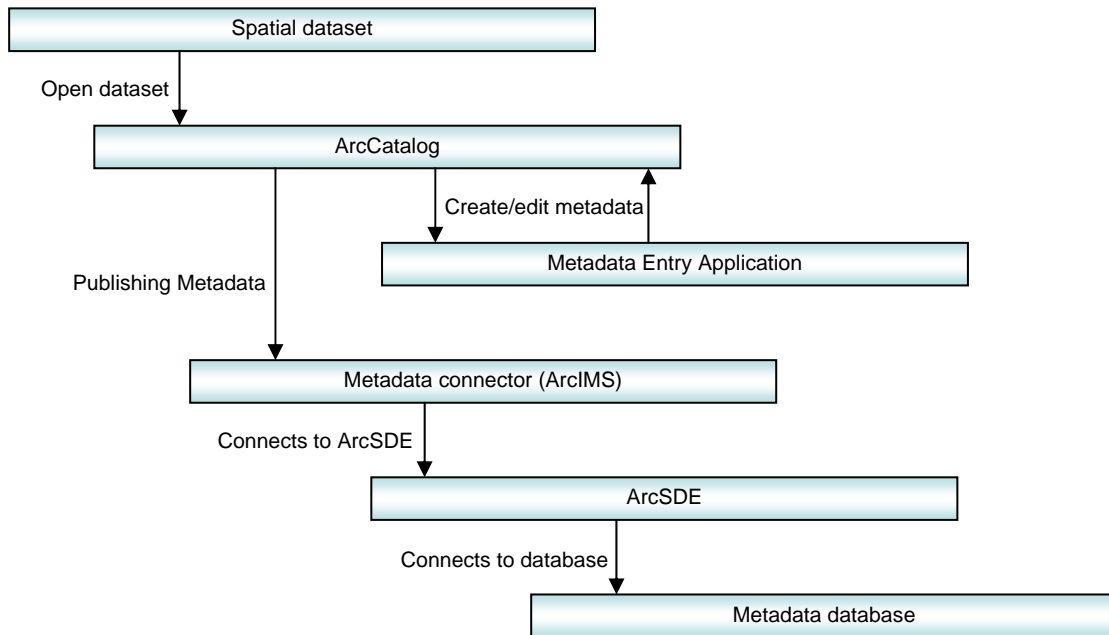


Figure 3: FSIC publish metadata work flow

Data management procedures

In order to make the catalog work properly, certain procedures need to be established. We identified three categories. Firstly procedures related to ownership; "who creates the meta-information", secondly procedures related to the quality control; "who checks the integrity of the meta-information" and thirdly, procedures related to the administration of the metadata service.

Preliminary results

During the implementation phase of the project, it was apparent that there were a number of pitfalls that needed to be addressed. One was that the catalog was not used as much as was anticipated, which might be related to the fact that it is a new service and people need to get used to it. Another was the risk of low quality of (meta-) data. During the cataloging process we came about data which contained all kinds of errors and in some cases we simply got the wrong metadata.

Reviewing the objectives related to the dissemination of the service within CIFOR, it became apparent that most people just don't get excited by metadata. They gave the metadata creation a low priority and they experienced the central data storage procedure a collective annoyance.

Conclusions

Put the user in the driving seat when it comes to implementing a (geo-) information catalog service. By doing so, you create demand, you (sometimes) simplify the design and it keeps the content focused. Promoting metadata development is necessary, though not easy to achieve, due to its unpopularity.

Quality control is a vital step in the data storage procedures, especially when it comes to metadata. Describing what is missing, unknown or wrong about a dataset, proves to be more valuable than having no information at all.

The following points conclude certain items that we feel need to be in place in order to successfully implement an information service:

- Constituency: identify your clientele;
- Statutes and relevance; clearly define the need for your service;
- Institutional capacity and stability; don't start anything beyond your capacity;
- Long-term funding;
- Acceptance, reliance and access;
- Evolutionary approach; use iterations; grow the service, don't build.

Acknowledgements

Many thanks goes to the CIFOR GIS Unit team members, who helped developing this system as well as adding data to it and helped promoting it within CIFOR and with their partners; Atie Puntodewo, Andy Darmawan, Miriam Van Heist, Iwan Kurniawan and Benoit Mertens. Special thanks go to Riza Nugraha for designing and developing the FSIC web service user interface.

Literature

Van Heist M. (Feb, 2002), Spatial Information Management at CIFOR.
Javanex (April 2002), Consultancy report on CIFOR data inventory
Legal Issues in the Use of Geospatial Data and Tools for Agriculture and Natural Resource Management

Federal Geographic Data Committee (www.fgdc.gov)
OpenGIS consortium (www.opengis.org)
The Apache Project (www.apache.org)
ESRI (www.esri.com)

Visit FSIC at <http://gislab.cifor.cgiar.org/fsic>

Joris Siermann
Systems architect FSIC
Information Services Group
Center for International Forestry Research (CIFOR)
Mailing address: P.O. Box 6596 JKPWB
Jakarta 10065, INDONESIA
Tel: (62-251) 622 622
Fax: (62-251) 622 100
E-mail: J.Siermann@cgiar.org