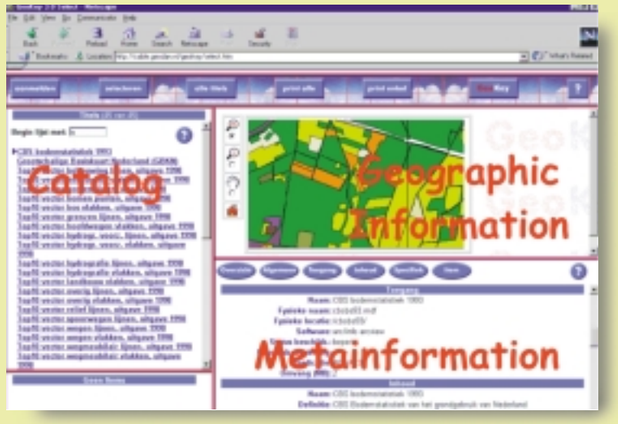
Searching and Sharing Spatial Information

The OpenGIS Catalog Specification Explained



Meta-information... Not so many years ago nobody in the GIS world did know what you were talking about when you mentioned the word "meta-information". Now, in 2000, there are national clearing houses for geo-information, there is a global spatial information infrastructure, there are new products like ESRI's ArcCatalog and there is a new standardization specification i.e. the OpenGIS Catalog Specification. This article describes the concept of meta-information, its catalog and explains the recent OpenGIS Catalog Specification. By Barend Gehrels, Technical Editor GeoInformatics

About six years ago, many organisations realized that over the past years they had collected many geographical datasets. They wanted to re-use more of these datasets. Therefore they described and ordered them for easy access and for use in their other projects. The description of a dataset is called meta-information. The collection of meta-information records is called a catalog. It has been quite an effort for those organisations to build up their meta-information catalogs, but it proved its value. They could re-use many datasets just because they did know of the existence of the data.

The Concept of the Clearinghouse

More recently, organisations wanted to access the meta-information of other organisations, as they wanted to mutually share their meta-information. This gave birth to the concept of the clearinghouse, an institution where meta-information of several organisations, or even a whole country, is collected and can be accessed. Users can have access to this clearinghouse to find datasets that satisfy their needs. They can do this via immediate retrieval of the datasets or by ordering them. However, connecting to these meta-information systems is not as simple as it looks. If meta-information systems are mutually dissimilar, connecting to them can be very difficult. To overcome this problem it would be of help to introduce standarization.

The description of a dataset is called meta-information. The collection of meta-information records is called a catalog

Meta-information

The Greek term meta denotes "a nature of a higher order or more fundamental kind". Meta-information means "information about information". Meta-information within the context of GIS describes information which by itself is geographic information, i.e. information that describes the earth. Meta-information basically consists of separate fields, together forming meta-information records. Some of these fields are descriptive, for example the source of the meta-information (the organization). Other fields give information about when and how the geographic dataset was created, i.e. creation date, date of actualisation, source scale, author, etcetera. Some fields are extracted from the geographic information itself, such as bounding box (the coverage), or topology-type (if the dataset contains points, lines, polygons, grid cells) and the projection.

Stored in a Standard Way

Meta-information should be stored in a standard way. Only then datasets are described consistently and users get used to the meta-information and know how to use it. In case of combining catalogs standardization becomes even more important. In Europe, the CEN (European Committee on Standards) standard is widely used. World-wide, the ISO (International Standards Organization) has defined a standard (ISO TC211), which is currently available in draft format. The OpenGIS Consortium embraces the ISO standard and will therefore not define its own standard for meta-information.

Catalog

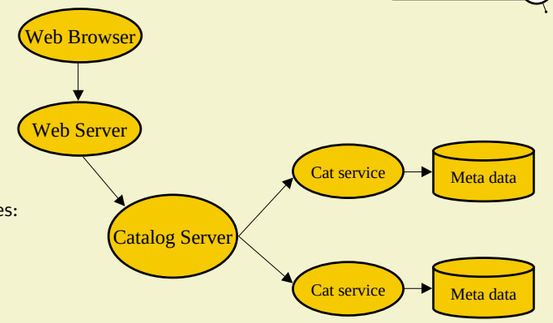
The word catalog is also a Greek word, meaning an ordered list or enumeration of objects. These objects can be books (catalogs are found in each library) but can also be, for example, stamps (in philately). Within the context of GIS the objects can be meta-information records. Thus, a collection of meta-information records about datasets can be set up as a catalog. Users can search through this catalog to find their metainformation records. Of course, in digital systems one can enter a query or criterion, and thus create a selection of records in which one is particularly interested. A meta-information system usually consists of the meta-information itself (the descriptive records) and the catalog function so as to let users find their meta-information. Sometimes there are links to the spatial information behind the meta-information in which case users can have access to this information. Some systems go even further and store a copy of the geographic information into the meta-information system. These systems are also called data warehouses because they contain and give access to everything: the meta-information, the geographic information itself and a method to find and get the data.

Access

Of course the current tendency is to give all users access to meta-information and catalog via the Internet and Intranet. Therefore, some kind of client/server system should be available. Catalogs are connected and must talk to each other by a common "language" (an interface or protocol). The Z39.50 specification is such an interface. It has its origin in the library world (a world which is abundant with catalogs) and is now also widely used in the geographic community. The American clearinghouse (FGDC) makes use of this protocol and also the ESMI catalogs in Europe are connected by a Z39.50 interface.

The OpenGIS Catalog Service

The OpenGIS Consortium (OGC) has been creating standards on GIS since 1994. They specified the so-called Simple Features, geographic entities that can be sent over a network in a uniform way. Until recently there were no OpenGIS standards for catalogs. In 1999, there was a proposal for a Catalog Interface specification which was accepted by OGC in August 1999 in Southampton.



The Catalog Interface consists of three services:
- discovery service
- access service
- management service

-Dicovery service
The discovery service enables the users to discover datasets. One can ask the discovery service questions in the form of queries, in which one can specify, for example, the title or a part of the title of the dataset, or the area the dataset must cover. Queries are specified in SQL (Standard Query Language), which is extended by OGC to process geographic queries with terms like INSIDE.

-Access service
The access service makes one retrieve or order geographic data. If one has found a dataset in the discovery service, one probably wants to view that dataset or to retrieve it for further use. Then the access service can be used for retrieval or processing orders.

-Management service
The management service enables users to add meta-information records on geographic datasets to a catalog. Also, a record can be modified or removed.

Meta-information should be stored in a standard way. Only then datasets are described consistently

-Catalog service
The catalog service enables distributed searches. Catalogs can be coupled to each other to form a network or a hierarchy. Users can send the query to one catalog. This catalog redirects the query to all connected catalogs. Then the catalog collects the results of al those catalogs and presents them to the user. The catalog service has various profiles for different Distributed Computing Platforms (DCP). There is a catalog service profile for CORBA, one for COM/OLE DB and one for the web. These profiles enable organisations to use the technology they already available. However, if a catalog of one profile must be coupled to a catalog of another profile, a bridge between those two catalogs is necessary.

The coming of these specifications has brought great potentials in the mutual network of catalogs. All catalogs of geographic metadata could be linked to each other in an Internet-style network.

Summary

Catalog Services supply users with a search program for spatial information. Since a catalog is filled with a list of meta-information records while catalogs are spread through various organisations, users can visit them to search through their catalogs for the desired information. To make this process more convenient to users, catalogs can be connected to each other to form together a spatial clearinghouse. To be able to connect catalogs to each other, standards are needed. These standards are supplied by the OpenGIS Consortium (OGC) and were accepted in August 1999 in the Southampton meeting. They will be open to the public in the near future. Finding spatial information will then be more convenient than it is now and clearing houses will be connected to each other so as to form the Global Spatial Information Infrastructure.

Barend Gehrels (barend@geodan.nl) is senior GIS consultant and developer at Geodan (the Netherlands) since 1994. He is technical representative at the Open GIS Consortium.

Catalog Proposal Team
(summer 1999)
Blue Angel Technologies, ESRI, Geomatics Canada, Intergraph, Marconi, MITRE, Oracle, FGDC, NASA, NIMA, Compusult, Geodan, HJW, JRC - European Commission, SICAD GEOMATICS