

Interoperable Spatial Data Catalogs

Douglas Nebert

Abstract

The discovery of spatial information on the Web is challenged by the many approaches to organizing and presenting it for search. Conventions for searching or navigating collections of information on the Web are almost as personal and unique as the Web authors who create them. Because there are few rules in the creation of Internet information sites, and because the number of these sites is increasing exponentially, the discovery of specialized, highly structured information — such as spatial information — is especially difficult. Communities that require or generate imagery, cartographic, and thematic spatial data usually manage their information in highly structured ways, but the organizing structures are not standardized. Inventories whose descriptive content is dynamic are often stored in a database, opaque to systematic web crawler searches. The use of full-text search engines and, even more recently, “meta” tags of keywords inside HTML (HyperText Markup Language) documents often falls short of imparting the contents of spatial data collections.

The OpenGIS Consortium (OGC) is in the process of defining implementable, interoperable spatial data catalogs that can be used to discover spatial data holdings in different data computing environments and across and within information communities. Spatial data catalogs have been around for many years, exemplified by product-specific image catalogs managed by remote sensing organizations. Unfortunately, there is little interoperability among spatial data catalogs, requiring the user or agent to traverse and translate many different user interfaces to locate relevant spatial data. In a broadly interoperable environment, a network of generic clients and servers could be built to enable global discovery of spatial data. As high-level software interfaces are standardized across disparate spatial data collections, access to them may be provided through reasonably lightweight gateway software, building up super-collections across and within information communities defined by discipline, geography, or cross-discipline interests.

Catalog Services

The OpenGIS Consortium has taken steps toward consensus on access to vector spatial information through its OpenGIS Simple Features Specification published in August 1997. The Simple Features Specification details the query of, and access to, simple geometric features (e.g., points, lines, areas) and basic attributes such that they can be used in external — often remote — applications directly. Access to Simple Features information is provided in the COM, CORBA, and SQL computing environments, and is aligned in approach through a single abstract Feature model. Similar specifications for “coverage” or raster-field types of data are being solicited at present to provide interoperable access to non-vector spatial data. With the addition of Catalog Services to the gallery of available interoperability specifications in 1999, the generalized ability to discover, access, and manipulate distributed

spatial data resources on the Internet and within Intranets through consistent methods will be realized. This is an essential precursor to the development of open distributed geoprocessing systems of the future.

OGC's March, 1998 Catalog Services Request for Proposals (RFP) was responded to by two submission teams: The Earth Observation and Geospatial (EO/GEO) Team and the Distributed Catalog Services (DCS) Team. Sponsoring companies involved in responding to the catalog services proposal include Oracle, ESRI, Blue Angel Technologies, and Intergraph, with contributions from the European Commission Joint Research Centre; Marconi Integrated Systems; Space Imaging; Star Technologies; Hammon, Jensen, Wallen and Associates (HJW); Lockheed-Martin; and the MITRE Corp. Governmental sponsors include NASA, the U.S. National Imagery and Mapping Agency (NIMA), the Canada Centre for Remote Sensing, and the U.S. Federal Geographic Data Committee (FGDC). Individuals participating in the two teams contribute considerable interest and expertise in remote sensing and GIS information management in the civilian, defense, and commercial sectors.

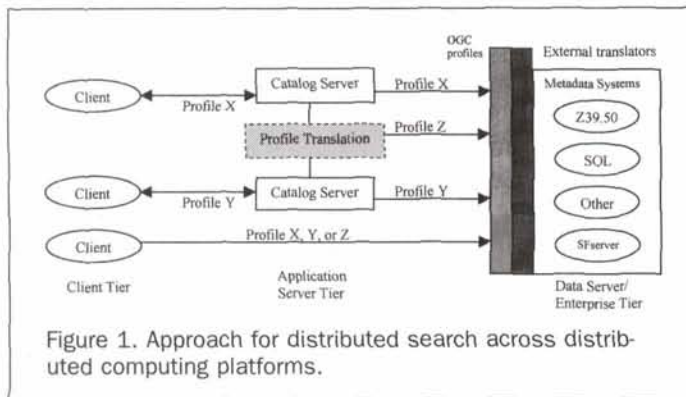
In the interest of providing a single, compatible solution for the OGC, the two teams are preparing a consolidated proposal that combines and extracts common functionality required by the sponsoring constituent communities. The DCS initial submission detailed a CORBA/COM functionality for a detailed enterprise image library capability, whereas the EO/GEO submission suggested a more protocol-independent approach that could be implemented in a variety of operational environments, including the Web, through the use of the Extensible Markup Language (XML) on HTTP. The two teams are now in the process of defining equivalence between the object approach of CORBA and the message-like approach of XML on HTTP so that interoperability bridges may easily be built and demonstrated between DCP-bound communities. If search interoperability were only defined on a single distributed computing platform (DCP) such as CORBA, or only within a single information community, then the notion of “global” catalog services for spatial data discovery and access could not be achieved. By supporting a higher level of potential compatibility for information discovery, interoperable solutions can be more readily built that satisfy both internal and external access interests.

Figure 1 illustrates the generalized approach being taken by the consolidation team to support requirements for distributed search across distributed computing platforms. Profiles of Catalog Service interfaces, labeled X, Y, Z, are shown as DCP-specific implementation specifications such as CORBA, COM, and XML-on-HTTP. A client using a given DCP-specific profile should be able to connect to an application server (catalog server) that may either (1) pass the query through to

Photogrammetric Engineering & Remote Sensing,
Vol. 65, No. 5, May 1999, pp. 573–575.

Clearinghouse Coordinator, Federal Geographic Data Committee Secretariat, USGS 590 National Center, Reston, VA 20192 (ddnebert@usgs.gov).

0099-1112/99/6505-573\$3.00/0
© 1999 American Society for Photogrammetry
and Remote Sensing



be responded to by one or more metadata systems or (2) link to profile translation services to gain access to a more heterogeneous collection of information services. Individual systems are accessed through translation software built upon standard "stubs" of code defined by the standard but linked to the local environment.

The scope of the combined submission is to define mandatory and optional capabilities of Catalog Services, and their corresponding messages, methods, and structures for each proposed DCP. These elements will be demonstrated in prototype implementations using adapted existing software in mid 1999 to show the feasibility of the approach. Conformance tests will be built at a later date to allow software to be tested for labeling as an "OpenGIS" conformant product — a measure of its stated compatibility.

There are many operational platforms on which Catalogs will be deployed, but the Web is the driving platform technology because of its potential for scalable wide-area computing and ability to integrate info-communities, etc. The submissions must handle different levels of capability (e.g., discovery, access, manage). These capabilities must have well-defined interfaces, and the interfaces must be made to interoperate within and between operational environments (distributed computing platforms, or DCPs).

What Does Metadata Mean?

There are competing requirements for catalog services with respect to the vocabulary of the metadata fields being managed. Within an enterprise or an organization, there may be a locally known set of fields for metadata query and presentation. In a broader community requiring access, there are usually "public" metadata fields to support cross-inventory or cross-discipline search that are advertised within the community. On a grander scale, there are likely to be some generic metadata properties that could describe almost all spatial data sets and aid in their discovery. In fact, without such a common set of searchable field meanings (semantics), broader search would be impossible.

The requirements of the sponsoring communities include the ability (1) to reference well-known or "public" metadata fields or (2) to allow private metadata fields where broader search interoperability is not required. The ISO Technical Committee 211 is drafting an international standard for over 200 public metadata element definitions and structures. The proposed solution in catalog services would be to qualify a given service as to the metadata model it uses and to reference registered semantics where they exist so that one can use or investigate the meaning of a given field in one or more languages or systems. Reference to such metadata dictionaries will greatly improve the ability to perform cross-discipline and cross-inventory searches on similar properties. Communities requiring a private schema will be able to ref-

erence their schema as an alternative, or possibly provide semantic mappings to support search using the private and public field tags.

The goal of the OpenGIS Catalog Services Specification is to, at a minimum, support discovery (perhaps also access and management) of geoinformation resources in a data-model independent way. Catalog interfaces should look and behave the same at a high level of abstraction no matter the structure and content of the metadata.

Discovery of Catalog Services

Implementation of Catalog Services within an organization or a small community may be enabled through a static list of available spatial data catalogs that can be searched. Within the Geospatial Data Clearinghouse activity sponsored by the FGDC, distributed web-based search access is provided to over 100 different collections or catalogs of spatial data worldwide, with additions occurring on a daily to weekly basis (see URL: <http://clearinghouse.fgdc.gov>). As the interest and availability of spatial data on the Internet increases, the ability to keep track of and navigate diverse catalog resources becomes accordingly complex. The growing interest in cross-disciplinary spatial data analysis in decision support and hazards information systems requires that more than one traditional discipline or custodian of data be visited, but such multi-site database searches are not quickly conducted.

Although the OGC Catalog Services proposal does not specifically address the discovery of spatial data resources on a global scale, international activities such as the Global Spatial Data Infrastructure (GSDI) effort would greatly benefit from an orderly means to discover and access all public spatial data catalogs and their holdings. The use of Catalog Services principles and software interfaces could be adopted and implemented as part of GSDI initiatives to provide global access points to register and search for spatial data at all levels of detail. A virtual global spatial data catalog composed of hundreds or thousands of online catalogs could be the result of this effort, linking suppliers and users of data, and seeding projects like Digital Earth with much needed descriptive information. The availability of commercial implementations of Catalog Services, the intended result of the OGC standards efforts, will help to enable the GSDI.

Bridging Information Communities and Protocols

The interoperable nature of OpenGIS-based technology can be interpreted at several different levels, yielding several different outcomes for success. At its simplest level, interoperability may mean that software components developed for and used in a single, homogeneous operating environment (for example, Microsoft Windows) may have predictable interfaces, which potentially result in solutions to a specific function being offered by multiple vendors. At this level of interoperability, however, applications built with such components may not be interoperable between heterogeneous computing platforms — the benefits of interoperability may stay at the programming level within a single platform and a single enterprise (or community). In a more systems-oriented view of interoperability, the component functions of OpenGIS specifications are expressed in stand-alone services, servers, and clients whose interfaces can be predictably accessed inside and outside the enterprise. In this domain, interoperability might be expressed in a flourishing environment of user and provider software, similar to that seen in Web servers and browsers today.

OpenGIS Catalog Services have the potential to be an important enabling technology for data discovery and sharing among information communities. Catalogs can play an important role in bridging the technical and institutional barriers that inhibit ubiquitous access to information. For catalog

services to flourish and be used in a broadly interoperable framework, the OpenGIS Catalog Services Specification must support familiar and popular protocols, both standard and community-based metadata content, and the ability to be implemented as "catalogs of catalogs" in a scalable fashion to permit wide and timely information discovery. To be eco-

nomically sustainable, catalog software components must be available in the commercial software marketplace. The current OpenGIS Catalog Services Specification effort takes this approach and should produce useful, interoperable, and commercially available solutions for spatial data discovery and access within the year.

Forthcoming Articles

Francisco Abarca-Hernández and Mario Chica-Olmo, Evaluation of Geostatistical Measures of Radiometric Spatial Variability for Lithologic Discrimination in Landsat TM Images.

M.M. Avar, F.R. Schiebe, and J.H. Everitt, Quantification of Chlorophyll in Reservoirs of the Little Washita River Watershed Using Airborne Video.

James R. Carr and Korblaah Matanawi, Correspondence Analysis for Principal Components Transformation of Multispectral and Hyperspectral Digital Images.

Albert K. Chong, A Technique for Spatial Sampling and Error Reporting for Image Map Bases.

Chang-fo F. Chung and Andrea G. Fabbri, Probabilistic Prediction Models for Landslide Hazard Mapping.

Nicholas Coops, Improvement in Predicting Stand Growth of *Pinus radiata* (D. Don) Across Landscapes Using NOAA AVHRR and Landsat MSS Imagery combined with a Forest Growth Process Model (3-PGS).

M.K. Crombie, R.R. Gillies, R.E. Arvidson, P. Brookmeyer, G.J. Weil, M. Sultan, and M. Harb, An Application of Remotely Derived Climatological Fields for Risk Assessment of Vector-Borne Diseases: A Spatial Study of Filariasis Prevalence in the Nile Delta, Egypt.

Xiaolong Dai and Siamak Khorram, Remotely Sensed Change Detection Based on Artificial Neural Networks.

F.M. Danson, N.A. Higgins, and N.M. Trodd, Measuring Land Surface Directional Reflectance with the Along-Track Scanning Radiometer.

Lesley-Ann Dupigny-Giroux and John E. Lewis, A Moisture Index for Surface Characterization over a Semiarid Area.

Rüdiger Gens, SAR Interferometry: Software, Data Format, and Data Quality.

Ayman F. Habib, Matching Road Edges in Stereo-Image Sequences Using Data Association Techniques.

Jung-Sheng Hsia and Ian Newton, A Method for the Automated Production of Digital Terrain Models Using a Combination of Feature Points, Grid Points, and Filling Back Points.

Y. Huang and J.C. Trinder, Object Recognition Based on Boundary Description.

W.B. Krabill, C.W. Wright, R.N. Swift, E.B. Frederick, S.S. Manizade, J.K. Yungel, C.F. Martin, J.G. Sonntag, Mark Duffy, William Hulslander, and John C. Brock, Airborne Laser Mapping of Assateague National Seashore Beach.

Lalit Kumar and Andrew K. Skidmore, Radiation-Vegetation Relationships in a Eucalyptus Forest.

Rongxing Li, Weian Wang, and Hong-Zeng Tseng, Detection and Location of Objects from Mobile Mapping Image Sequences by Hopfield Neural Networks.

Donald L. Light, C-Factor for Softcopy Photogrammetry.

Paul D. Lowman, Jr., Landsat and Apollo: The Forgotten Legacy.

Ross S. Lunetta and Mary E. Balogh, Application of Multi-Temporal Landsat 5 TM Imagery for Wetland Identification.

Emilio F. Moran, Remote Sensing and GIS at Farm Property Level: Demography and Deforestation in the Brazilian Amazon.

Thierry Ranchin and Lucien Wald, Fusion of High Spatial and Spectral Resolution Images: The ARSIS Concept and Its Implementation.

John P. Rowe, Timothy A. Warner, Darrell R. Dean, Jr., and Andrew F. Egan, A Remote Sensing Strategy for Measuring Logging Road System Length from Small Format Aerial Photography.

Michael J. Schuft, Thomas J. Moser, P.J. Wigington, Jr., Don L. Stevens, Jr., Lynne S. McAllister, Shannen S. Chapman, and Ted L. Ernst, Development of Landscape Metrics for Characterizing Ribarian-Stream Networks.

Youngsinn Sohn, Emilio F. Moran, and Francisco D. Gurri, Deforestation in North-Central Yucatan (1985-1995): Mapping Secondary Succession of Forest and Agricultural Land Use in Sotuta Using the Cosine of the Angle Concept.

Benoît St-Onge, Topographic Effects on the Texture of High-Resolution Forest-Stand Images Measured by the Semivariogram.

Guoqing Sun, K. Jon Ranson, Jack Bufton, and Michael Roth, Requirement of Ground Tie Points for InSAR DEM Generation.

D.D. Sworder, J.E. Boyd, G.A. Clapp, and R. Vojak, Nonlinear Trackers Using Image-Dependent Gains.

A.B. Tait, D.K. Hall, J.L. Foster, and A.T.C. Chang, High Frequency Passive Microwave Radiometry over a Snow-Covered Surface in Alaska.

Jan Verhoeve and Robert De Wulf, An Image Processing Chain for Land-Cover Classification Using Multitemporal ERS-1 Data.

Jim Vrabel, Multispectral Imagery Advanced Band Sharpening Study.

Timothy G. Wade, James D. Wickham, and David F. Bradford, Accuracy of Road Density Estimates Derived from USGS DLG Data for Use in Environmental Applications.

Fangju Wang, Query Optimization for a Distributed Geographic Information System.

Jianjun Wang, Gary J. Robinson, and Kevin White, Generating Viewsheds without Using Sightlines.

Dean Whitman, Timothy Gubbels, and Linda Powell, Spatial Interrelationships between Lake Elevations, Water Tables, and Sinkhole Occurrence in Central Florida: A GIS Approach.

Stephen S. Young and Assaf Anyamba, Comparison of NOAA/NASA PAL and NOAA GVI Data for Vegetation Change Studies over China.

W. Zeitler, T. Ohlhof, and H. Ebner, Recomputation of the Global Mars Control Point Network.

Jingxiong Zhang and Roger P. Kirby, Alternative Criteria for Defining Fuzzy Boundaries Based on Fuzzy Classification of Aerial Photographs and Satellite Images.

Guoqing Zhou, Jeorg Albertz, and Klaus Gwinner, Extracting 3D Information Using Spatio-Temporal Analysis of Aerial Image Sequences.