Towards the construction of the Spanish National Geographic Information Infrastructure

J. Nogueras[1], M.A. Latre, M. Navas, R. Rioja, P.R. Muro-Medrano
Department of Computer Science and System Engineering
University of Zaragoza
María de Luna 3, 50015 Zaragoza (Spain)
[1]jnog@ebro.cps.unizar.es

Abstract. The creation of specific organisations or programs for developing or overseeing the development of Spatial Data Infrastructures, particularly by government at various levels, can be seen as the logical extension of a long practice police oriented to co-ordinate the building of other necessary infrastructures for ongoing development such as transportation or telecommunication networks. This kind of infrastructure is already under construction in many nations within our technological sphere but yet noticeably absent in Spain. Recently, researchers from three universities in Spain have launched a project with the intention to encourage the development of a National Geographic Information Infrastructure in Spain. This work presents some functionality aspects of the system to be built for this project. This functionality is going to be supported by three main technological groups of components: Metadata catalogs, Web mapping servers, and Geographical data access and e-commerce.

1. Introduction

According to [Nebert], the term Spatial Data Infrastructure (SDI) is often used to denote the relevant base collection of technologies, policies and institutional arrangements that facilitates the availability of an access to spatial data. The SDI provides a basis for spatial data discovery, evaluation, and application for users and providers within all levels of government, the commercial sector, the non-profit sector, academia and by the average person.

The word infrastructure is used to promote the concept of a reliable supporting environment, analogous to a road or telecommunications network, that, in this case, facilitates the access to geographically-related information using a minimum set of standard practices, protocols, and specifications. The applications that run on such an infrastructure are not specified in detail in this document. But, like roads and wires, a SDI facilitates the conveyance of virtually unlimited packages of geographic information.

A SDI must be more than a single data set or database; an SDI includes geographical data and attributes, enough documentation (metadata), a means to discover, visualize, and evaluate the data (catalogues and Web mapping), and some method to provide access to the geographical data. Beyond this, there are additional services or software to support applications of the data. To make an SDI functional, it also must include the organizational agreements needed to coordinate and administer it on a local, regional, national, and or trans-national scale.

The creation of specific organisations or programs for developing or overseeing the development of SDI, particularly by government at various scales can be seen as the logical extension of the long practice of co-ordinating the building of other infrastructures necessary for ongoing development, such as transportation or telecommunication networks. Spain is wholly absent from comparative SDI studies such as [Masser et al 1998], [Masser 1999] and [Craglia et al] because it has built no such infrastructure. This may surprise the reader given the nation's size and relative importance within Europe, also given that neighbouring Portugal does have a relatively mature SDI.

Recently, researchers from three universities in Spain have launched a project with the intention to encourage the development of a National Geographic Information Infrastructure in Spain. This work presents some functionality aspects of the system to be built for this project. This functionality is going to be supported by three main technological groups of components:

- Metadata catalogs. Large amounts of geospatial data, which are stored for use in local databases, can often be accessed from external applications once they are published. Geospatial data catalogs are the solution to publish descriptions of such geospatial data holdings in a standard way that enables search across multiple servers. Geospatial data catalogs are discovery and access systems that use metadata as the target for query on raster, vector, and tabular geospatial information. The use of indexed and searchable metadata provides a selected and disciplined vocabulary against which intelligent geospatial search can be performed within or among geographical infrastructure user communities.
- Web mapping servers. Historically, the primary visualization of geographical data has been achieved through maps. In the context of a geographical infrastructure, it is increasingly useful to provide mapped or graphical views of geospatial data through online mapping interfaces. This way, it is possible to evaluate data and satisfy many of the needs of novice or browse users without requiring the full data download. Although it is not a replacement for direct data access, it satisfies a broad requirement for public interaction with geospatial information. Assuming that data are being used for their correct purpose and on an appropriate scale, maps can quickly portray a large amount of information to the inquirer. The rise of the Internet and in particular the World Wide Web has allowed information providers to harness this technology to the conventional stove-pipe GIS systems and data warehouses. The work to be done for this group of components will be guided by the OpenGIS Consortium's Web Mapping Specification.
- Geographical data access and e-commerce. Once spatial data of interest have been located and evaluated, using the Catalog and online mapping techniques previously described, access to detailed geospatial data in its packaged form is often required by advanced users or application software. Access involves the order, packaging and delivery, offline or online, of the data (coordinate and attributes according to the form of the data) specified. Finally, exploitation is what the consumer does with the data for their own purpose. The work to be done will be focused in offering services to provide this access functionality. In many cases, this access will involve the necessity of economic transactions. Therefore, this group of framework components will then open the door to possible geographical data e-commerce.

All these groups of components and services (Web Mapping Services, Catalog Services) are based on new and maturing technologies. Their own evolution will provide us with additional services to be identified and selected. Moreover, the particular circumstances and requirements of partners of this project (companies and public administrations) will introduce new problems, policies and restrictions, which should be evaluated and if necessary incorporated to the National Infrastructure to be built.

The rest of the paper is structured as follows. Next section presents catalog components. Section three shows the Web mapping services. Section four talks about the geographical data access and e-commerce. This work ends with a conclusion section.

2. Catalog components

From a broader perspective, a catalog cannot be understood as a single element implemented like a simple server running on a high-availability host connected to the Internet. It is a set of components developed to provide functionality to create metadata, to store this metadata into a special repository, and to search on it. Figure 1 presents these components and their relationships.

There are four basic components. The first one is the *CatMdEdit*. It is the component which enables the different agents of a spatial data infrastructure to create consistent metadata describing geospatial data resources. This component has two functionality levels. The lower one provides the basic services for creating and editing metadata. Apart from this basic functionality, the component is optionally enhanced with a set of tools to improve the quality of metadata: a thesaurus management tool and a metadata automatic generation tool. The first tool enables metadata originators to use look-up tables in order to fill some metadata elements with pre-defined lists of controlled keywords. The use of these controlled keywords facilitates the mapping between a selected vocabulary and a large collection of datasets. This way, the catalog search tools make possible the discovery of data based on hierarchies of concepts. The thesaurus management tool integrates a relational database storing classifications provided by recognized authorities (some examples are found in the NASA 2000 or ADL 2000). The second optional tool is an automatic metadata generation tool which is able to derive metadata from the data sources by means of interconnection with commercial GIS tools or proprietary software. Examples of derived metadata are information about spatial reference systems, number and type of geographic features, extension covered by a dataset, or information about the entities and attributes of alphanumerical related data.

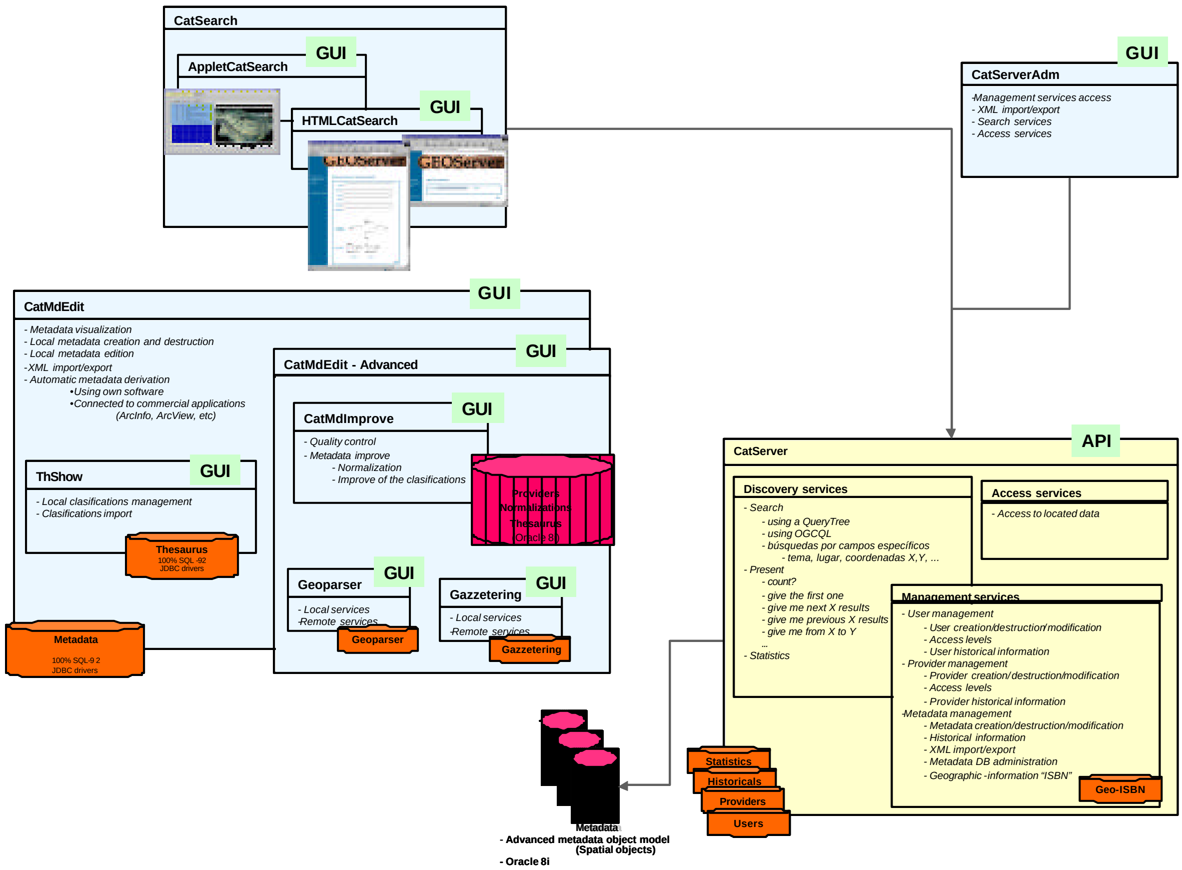


Figure 1: *Geographic catalog components.*

This lower level works with a "cheap" database (requires uniquely the access by means of JDBC to any Relational Data Base Management System such as Access, Oracle, mySQL, …) which is responsible for the storage of the metadata entries using a SQL-92 metadata database model. The upper level of functionality is oriented to advanced metadata contributors, as well as to catalog administrators in charge of management and improvement of metadata controlled under the geographic catalog. In this case, the system works with Oracle 8i because of its special capacities for the use of spatial objects, text management (Oracle IntermediaText) and thesaurus.

Details about the metadata supported by the *CatMdEdit* (and the other components presented in this paper), its GUI and special features can be found in [Zarazaga et al].

The *CatSearch* component groups two generic types of search tools. The first one (*AppCatSearch*) is an applet search tool which provides users with discovery capabilities on a Intranet/Internet environment interacting with *CatServer*. This tool provides as well different interfaces to construct the query expressions which are processed by *CatServer*. A generic interface enables the construction of expressions composed of unlimited restrictions on metadata element values. A second set of interfaces enables a "free-text" (natural language approach) query construction; and "map-guided" interface restricting the area of interest of the desired data sources to discover. In addition to this, it is also possible to perform searches based on available keyword classifications. Finally, the second search tool is a light-weight HTML client oriented to Internet clients which provides public access to contents of the *CatServer* by means of a simple interface.

The third component is the *CatServerAdm*. It is the GUI component which permits the complete administration of the *CatServer*. These administration services include user management, provider management or load balance administration of catalog local storage. It also incorporates functionality for searching similar to the previous components.

The last one, and maybe the most relevant one, has been named *CatServer*. This component is a server whose development has been guided by OGC Catalog Services specification ([OGC 1999a]). The functionality specified by the interfaces of this specification has been taken into account for the development of *CatServer*. Furthermore, it is expected that in the future, *CatServer* will offer an interface compatible with the Coarse-Grain Structural Model and interoperable among different Distributed Computing Plattforms. However, with respect to this work, this server runs as a Java-RMI server ([Orfali et al]), which provides OGC services as well as additional services supporting catalog users and metadata providers (metadata contributors).

3. Web mapping services

Imagine you want to buy a car. The only information you have about it is the number of wheels, the colour, the make and the price (car-metadata), but you never have seen this model. You would like to see it, even a photo. Same problem happens with the geographical information. Consider a common situation in the real world. An organization is involved in the study of a certain geographical zone to make any kind of study with the geographical data associated. They have located a set of geographical information related with this area using the catalog. They have found some possibilities of maps and they must decide what to buy. Usually, map prices are not “very cheap”, so they have to be careful in their decision. In this case, it would be desirable to have a “photo” of the data to be able to evaluate the map. Web mapping services can offer this “photo” to costumers.

Serving maps on the Internet is a topic of research nowadays. The interest of users in publishing geographical information on the net, in addition to the ability to get and display information from different sources, have made businesses to make a special effort to cover the clients' requirements. The WMS proposed by OpenGIS and developed with the collaboration of several company answers to the exposed requirements. The interface specifies a set of services designed to allow the request of maps on the Internet. Commercial OpenGIS servers must implement this interface in order to enable clients, or even other servers, to interchange information. A work environment that needs daily updated geographical information is the best site to introduce a central repository of data accessible through a WMS. The possibility of accessing to several central repositories that offer a WMS interface makes it easier for users to work with local information combined with the last updated data from different catalogues.

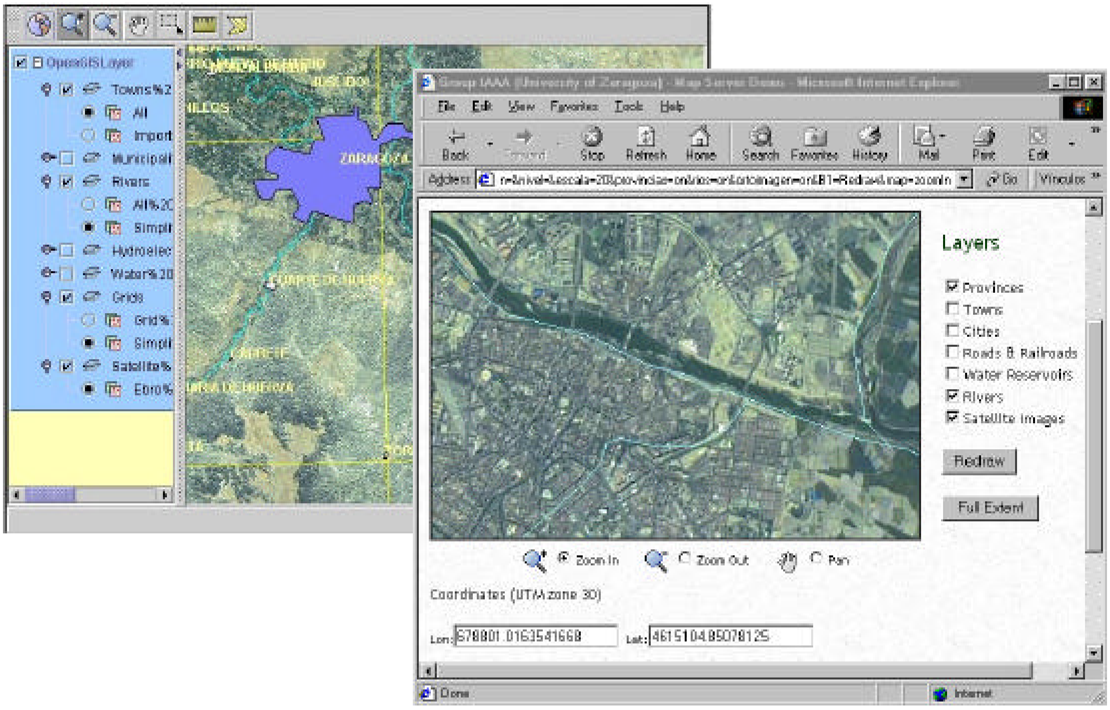


Figure 2*: Web map server clients.*

The OpenGIS specification of web interfaces, addressed to the most chaotic information source ever known, is a great step to allow the interoperation of multiple users of geographical information across the world. Central repositories of data with common interfaces are one of the main topics of OpenGIS. Clients will use WMS services to obtain information from a central repository, or even from several ones. Cooperation is only possible if the interfaces to access the resources are well known. The web map servers are designed to interoperate and to allow the sharing of geographical information among users. By example, a geographer from the mining department wants to display dynamically a map of a mine with satellite images of the institute of image processing, the level curves of the terrain he has, and the artesian wells from the hydrological department. If all the data sources have the information accessible through a web server offering the OpenGIS WMS interface, our geographer will be able to generate and navigate over the dynamically created map, without having to download the individual files with the data, and having always the latest version of the data.

Currently, there is available a Web map server in compliance with the OpenGIS Web Map Server Interface Specification version 0.9 ([OGC 1999b]). It has been made in Java and uses the geographical data management and rendering capabilities of a Java GIS-kernel to generate and display maps. It adds the functionality needed in order to fulfil the OpenGIS interfaces. It is in charge of receiving the map, capabilities or feature information requests, generating the answer as a file, and returning it to the requester. Figure 2 offers two kind of WMS clients (HTML and Java-applet) currently available. The system also offers a Java RMI interface with the same services described by the OpenGIS web map server (details about this WMS can be found in Fernandez et al).

4. Geographical data access and e-commerce

One time the geographical resources have been discovered and evaluated, next step is to access to them according with the paradigm proposed in [Nebert] (Figure 3). In this way, a SDI should offer tools for data providers in order to facilitate the management of the data, and to specify the geographical-information processes. These processes could include "on-line" and "off-line" access. Furthermore, providers should have mechanism to offer "free-access" data and "pay-access" data.

"Pay-access" data, "on-line" or "off-line", can represent the e-commerce view of the GI access. According to a study made by Cooper & Lybrand ([GI2000]), the GI production market in Europe has been estimated in about 10 Billions-Euro per year (in Spain this market has been estimated in only 70 Millions-Euro). But this data is being collected, not selling from providers to customers. This market area has not been enough exploited. Furthermore, current transactions of data used to be done "off-line", using external repositories such as CDs, and with complicated and expensive order, purchase and deliver processes. If GI can been located and evaluated through the Web, the acquisition process should be able to be done throw Internet as well.

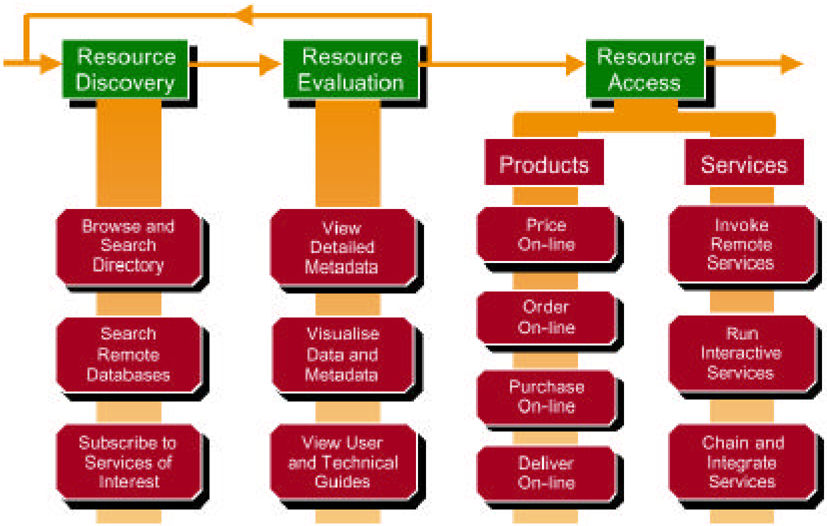


Figure 3: *Geospatial Resource Access Paradigm.*

The technological support of these access processes can be developed with very different approximations. For example, the simple one could be implemented like URLs where the whole data is accessible. The way to find these URLs could be guided by a set of Web pages and forms which could included the collection of information about credit cards or other pay methods. One example of "free-access" with URLs can be find in the Ebro River Basin Web page (http://www.chebro.es). Other way to access services can be done with the use of Web feature servers ([OGC 2001]). This approximation offers a higher degree of flexibility and services because allows the access to concrete geographical items such as rivers, roads, etc.

Currently, a relationship between our laboratory and one of the most important Spanish-companies in geographical information creation and management has been established in order to study, design and prototype components for GI e-commerce. This work will take into account the evolution of OGC specifications in this area.

5. Conclusions

This work has presented some functionality aspects that are been taken into account in the development of a National Geographic Information Infrastructure in Spain. This is a project that has been launched recently by researchers from three universities in Spain since this kind of infrastructure, already under construction in many nations within our technological sphere, is yet noticeably absent in Spain.

This functionality is going to be supported by three main technological groups of components: Metadata catalogs, Web mapping servers, and Geographical data access and e-commerce. The main kernel of these sets of components is going to be built using Java language and related technologies. Nevertheless, it will be necessary to use other languages, especially when commercial products and libraries will be connected to this kernel.

Acknowledgements

This work has been partially supported by the Spanish Interministerial Council on Science and Technology (CICYT), grant numbered TIC2000-1568-C03-01.

References

ADL 2000, *Alexandria Digital Library Project*, http://www.alexandria.ucsb.edu.

NASA 2000, *NASA Master Directory.Draft geospatial thematic keywords from the NASA Master Directory in short and long format for CSDGM of FGDC* ,http://www.fgdc.gov/clearinghouse/reference/refmat.html,2000

Nebert, D.D. (eds), *Developing Spatial Data Infrastructures: The SDI Cookbook v.1.0*, Global Spatial Data Infrastructure (http://www.gsdi.org), 6 July 2000.

Craglia, M., Annoni, A.,. Masser, I., *Data Policy In Europe: National and Regional Perspectives* , Luxembourg, European Commission , http://www.ec-gis.org/reports/policies.pdf, 2000

Masser, I., "All shapes and sizes: the first generation of national spatial data infrastructures", *Int Journal of GIS*, no. 13, pp. 67-84, 1999.

Masser, I., Burrough, P., "International aspects of spatial data exchange", *European geographic information infrastructures: opportunities and pitfalls*, pp. 1-14., P. Burrough and I. Masser (eds.), Taylor and Francis (London), 1998

OGC 1999a, *OpenGIS Project Document 99-051s, OpenGIS – Catalog Interface Implementation Specification (version 1.0)*, OpenGIS Consortium, 1999

OGC 1999b, *OpenGIS Project Document 99-077r4, "OpenGIS Web Map Server Interface Specification (version 0.9)"*, OpenGIS Consortium, 2000.

Orfali, R., Harkey, D., *Client/Server Programming with Java and CORBA*, 2nd Edition, Ed. Wiley Computer Publishing, 1998.

Fernández, P., Béjar , R., Latre, M.A., Valiño, J., Bañares, J.A., Muro-Medrano, P.R., "Web mapping interoperability in practice, a Java approach guided by the OpenGis Web Map Server Interface Specification", *EC-GIS 2000, 6th European Commission GI and GIS Workshop*, Lyon (France), 28–30 June 2000.

Zarazaga, F.J., López, R., Nogueras, J., Cantán, O., Álvarez, P., Muro-Medrano, P.R., "First Steps to Set Up Java Components for the OpenGIS Catalog Services and its Software Infrastructure", *3rd AGILE Conference on Geographic Information Science*, Helsinki/ Espoo (Finland) , 25–27 May 2000

GI2000, *"GI & GIS Market Size Data"*, European-Commission supported publications (DG XIII-E), http://158.169.50.95:10080/gi/en/markets/gimrktfigs.html.

OGC 2001, *"OpenGIS Web Feature Server Specification". Draft candidate implementation specification*, OpenGIS Consortium, 2001.

European Commission
SPE 01.150 – Proceedings of the 7th EC-GI & GIS Workshop 'EGII - Managing the Mosaic'

Editors: Karen Fullerton & Alison Munro

The 7th EC-GI & GIS Workshop was organised by the GI & GIS project of the Joint Research Centre. The GI & GIS project supports the actions to create an INfrastructure for SPatial InfoRmation in Europe (INSPIRE). INSPIRE encompasses the broad policy, organisational, technical and financial arrangements necessary to support increased access to Geographic Information in Europe. It will potentially benefit many stakeholders - government and non-government organisations, education and research institutions, the commercial sector, and the general community - at the national, regional and global level. INSPIRE offers the prospect of better decision-making and thus improved economic growth, social development and environmental management.

The GI & GIS project reflects the increasing awareness of the importance of spatial aspects in policy and decision-making, and focuses on both policy and technical issues relating to spatial information and spatial information systems. The main activities of the project include:

- Assisting the GI policy making process of the Commission, helping to formalise the user requirements for INSPIRE, and contributing to the establishment of an interoperable European Geo-Statistical system;

- Developing harmonised and coherent multidisciplinary Pan-European databases and analyzing spatial information across different sectoral policies and different levels of governmental organizations;

- Providing technical support to the Services of the Commission in defining base and thematic spatial data requirements at the European level, and operating communication facilities; and

- Monitoring GI & GIS standards, interoperability, and market development through a Technology Watch activity.

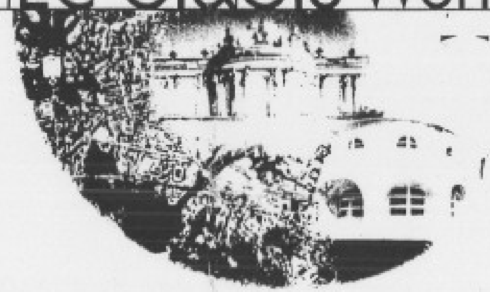
For more information on GI & GIS in the European Commission please visit our website: **http://www.ec-gis.org/**

7EC-GISWS

Proceedings

EGII
Managing the Mosaic

7th EC-GI&GIS Workshop



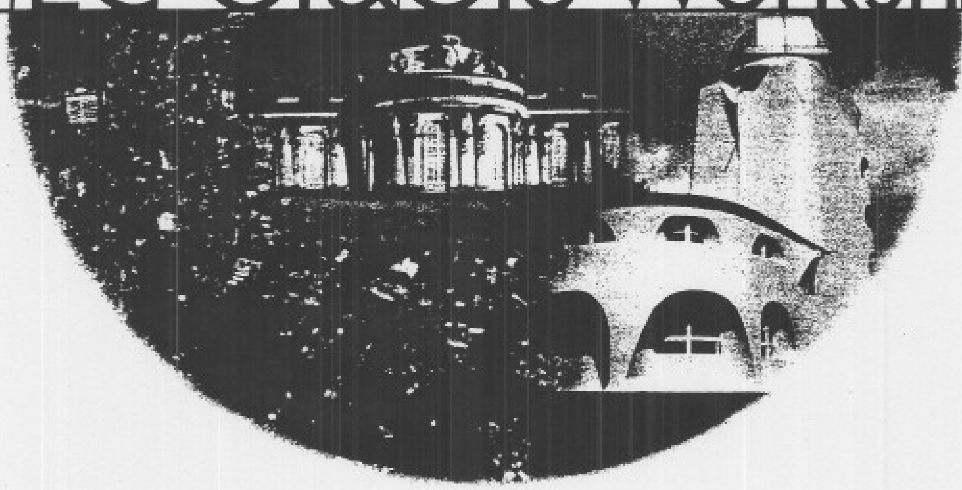
Potsdam, Germany
13-15 June 2001

Space Applications Institute

Abstracts

EGII
Managing the Mosaic

7th EC-GI&GIS Workshop



Potsdam, Germany
13-15 June 2001

Space Applications Institute

EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

GFZ
POTSDAM