Developing Spatial Data Infrastructures:

The SDI Cookbook

Version 1.1
15 May 2001

Editor: Douglas D. Nebert, Technical Working Group Chair, GSDI

Table of Contents

Chapter One: The Cookbook Approach

Introduction

At the United Nations Conference on Environment and Development in Rio de Janeiro in 1992, a major resolution was passed to focus on reversing the impacts caused by environmental deterioration. The Agenda 21 resolution establishes measures to address deforestation, pollution, depletion of fish stocks, and management of toxic wastes to name a few. The importance of geographic information to support decision-making and management of these growing national, regional, and global issues was cited as critical at the 1992 Rio Summit, and by a special session of the United Nations General Assembly assembled in 1997 to appraise the implementation of the Agenda 21.

Geographic information is vital to make sound decisions at the local, regional, and global levels. Crime management, business development, flood mitigation, environmental restoration, community land use assessments and disaster recovery are just a few examples of areas in which decision-makers are benefiting from geographic information, together with the associated infrastructures (i.e. Spatial Data Infrastructure or SDI) that support information discovery, access, and use of this information in the decision-making process.

However, information is an expensive resource, and for this reason appropriate information and the resources to fully utilise this information may not always be readily available, particularly in the developing world. Many national, regional, and international programs and projects are working to improve access to available spatial data, promote its reuse, and ensure that additional investment in spatial information collection and management results in an ever-growing, readily available and useable pool of spatial information. This is true of many initiatives even if they are not actually labelled as “SDI initiatives”. An example of this is the Environment Information System Program in sub-Saharan Africa (EIS-SSA). An emphasis on harmonising standards for spatial data capture and exchange, the co-ordination of data collection and maintenance activities and the use of common data sets by different agencies may also feature in such initiatives, although these activities by themselves do not constitute a formal SDI.

In regions characterised by an availability of geographic information, in combination with the power of Geographic Information Systems (GIS), decision support tools, data bases, and the World Wide Web and their associated interoperability, the way better-resourced communities address critical issues of social, environmental, and economic importance is changing rapidly. However, even in the new era of networked computers, the social habits of the past continue to prohibit users from finding and using critical geographic information. This can lead to either the abandoning of a proposed project, or to unnecessary – and expensive - recapture of existing geographic information.

There is a clear need, at all scales, to be able to access, integrate and use spatial data from disparate sources in guiding decision making. Our ability then, to make sound decisions collectively at the local, regional, and global levels, is dependent on the implementation of SDI that provides for compatibility across jurisdictions that promotes data access and use.

Only through common conventions and technical agreements will it be easily possible for local communities, nations and regional decision-makers to discover, acquire, exploit and share geographic information vital to the decision process. The use of common conventions and technical agreements also makes sound economic sense by limiting the cost involved in the integration of information from various sources, as well as eliminating the need for parallel and costly development of tools for discovering, exchanging and exploiting spatial data. The greater the limitation on available resources for SDI development, the greater the incentive for achieving alignment between initiatives to build SDI.

The development of a "cookbook" is envisaged as a means to clarify the SDI definition and to share the current experiences in building SDI implementations that are compatible at many scales of endeavour. This cookbook is intended to be a dynamic document available in printed and digital form, to include "recipes" or recommendations on developing these infrastructures from a local, even non-governmental, scale through global initiatives.

Scope of this Cookbook

This SDI Implementation Guide or Cookbook, through the support of the Global Spatial Data Infrastructure community, provides geographic information providers and users with the necessary background information to evaluate and implement existing components of SDI. It also facilitates participation within a growing (digital) geographic information community known as the Global Spatial Data Infrastructure (GSDI).

To enable builders of SDI to make use of and build on existing SDI components in a way which makes their endeavors compatible with the efforts of other SDI builders, this GSDI Cookbook identifies:

- existing and emerging standards,
- free and commercial standards-based software solutions,
- supportive organisational strategies and policies and
- best practices.

Working within a common framework of standards and tools based on these standards also makes it possible to maximise the impact of the total available resources for SDI creation through future co-operation -- e.g. we develop this, you develop that, and then we share.

Although proprietary or project-based solutions for information sharing continue to exist, the adoption of consistent geospatial data sharing principles will in general provide a better solution for information dissemination, through publishing geospatial data using the Internet and computer media. In an increasingly "global community", there is a need to ensure that trans-national implementations and common knowledge bases are available. Ultimately, these SDI activities should improve collaboration within the geospatial data industry and make the benefits derived from the use of geographic information part of everyday life for all.

Spatial Data Infrastructures

The term "Spatial Data Infrastructure" (SDI) is often used to denote the relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and 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 citizens in general.

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, an SDI facilitates the conveyance of virtually unlimited packages of geographic information.

An SDI must be more than a single data set or database; an SDI hosts geographic data and attributes, sufficient documentation (metadata), a means to discover, visualize, and evaluate the data (catalogueues and Web mapping), and some method to provide access to the geographic data. Beyond this are additional services or software to support applications of the data. To make an SDI functional, it must also include the organisational agreements needed to coordinate and adminster it on a local, regional, national, and or trans-national scale. Although the core SDI concept includes within its scope neither base data collection activities or myriad applications built upon it, the infrastructure provides the ideal environment to connect applications to data – influencing both data collection and applications construction through minimal appropriate standards and policies.

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.

The Global Spatial Data Infrastructure

Just as SDI programs of necessity involve the alignment of scarce resources for achieving success, so too it is necessary to ensure that the SDI initiatives develop in harmony with each other in order to maximise the impact of these programmes. In reality, many initiatives are working in isolation, not necessarily developing in harmony with others and consequently unable to reap the benefits of working together.

Anyone who is involved in a project of which spatial information forms an integral part and who intends leaving a legacy of spatial data or tools to exploit the data that lasts beyond the period of funding for the project is, by definition, participating in some of the fundamental elements required by an SDI. As coordination between such organisations expands, these projects very often lay the foundations on which initiatives formally dedicated to the establishment of SDI can then build. See Chapter 9 for specific case studies.

At a global scale, the most prominent examples of formal SDI programs are on a national scale. Most of these are driven by the national or federal government (e.g. the NSDI in the USA, the SNIG in Portugal, Australia's ASDI, Malaysia's NaLIS, South Africa's NSIF, Colombia), but there are exceptions such as the Uruguay Clearinghouse and NGDF in the United Kingdom, which have largely been driven by the private sector. In most cases the need for wide participation in the development of lasting, useful SDI is acknowledged, and so private-public partnerships are encouraged. The beneficiaries of SDI are generally seen to derive from the public and private sectors, academia and non-governmental organisations, as well as individuals. Federal countries are often able to build their national SDI programs on SDI programs being driven by provincial or state governments (e.g. the ASDI of Australia). Transnational SDI initiatives often arise out of existing transnational structures (e.g. the Permanent Committee for GIS Infrastructure in Asia and the Pacific was formed through the UN Regional Cartographic Conference for the Asia-Pacific region).

Distribution

This GSDI Cookbook is intended to be a "living" and dynamic document that can be updated as new principles and technologies are adopted. Distribution of this Cookbook is intended primarily via the World Wide Web, although electronic copies will also be made available on other physical media such as CD-ROM and printed copy for audiences that are not well connected to the Internet at this time.

Should you be reading this via the World Wide Web and wish to obtain a soft or hard copy, please contact the GSDI secretariat, at www.gsdi.org .

Contributors

Contributions to this GSDI Cookbook are indeed global and are intended to satisfy many different categories of participants. Contributors from around the world have nominated or been selected to organise and contribute to each chapter. This was a deliberate choice, in order to ensure that the Cookbook represented various perspectives from around the globe, to ensure both that the collective global experience and existing resources would be represented in the Cookbook, and that its applicability could truly be global.

Ongoing contributions to this GSDI Cookbook are welcome, and indeed necessary. If you believe that you have something to contribute to the cookbook, please contact the GSDI Technical Working Group through www.gsdi.org.

Organisation

Each chapter is organised into three major sections that correspond to levels of detail and application:

- The first section in each chapter establishes the background, context, and rationale for the subject suitable as general orientation for all readers, but targeted for managers and end-users
- The second section addresses the design architecture of organisations, roles, and software systems that are intended to interact
- The third section addresses the implementation with review of existing standards, protocols, and software as appropriate

Each chapter is approximately 10 to 20 pages in length with links to other relevant documents. Use-case scenarios and illustrations are featured in some chapters as inset boxes to further build understanding. Each chapter has a set of recommendations placed in a summary. Use of terminology is (will be) hyper-linked to Chapter 10 (Terminology) in the online version or is underlined as implied links to the terminology section in the printed version.

Case studies are intended to provide for local or regional relevance and interpretation. The document style not intended to be overly technical, however contributors have provided references to more detailed and comprehensive technical information where possible.

Finally, no manual of this type can claim to provide all the answers to all the possible national spatial data infrastructure permutations that exist. This cookbook does provide a basic set of guiding principles that have been successful for establishing compatible Spatial Data Infrastructures, and are supported by the Global Spatial Data Infrastructure to promote successful decision-making for issues of local, regional, and global significance. As mentioned in the preceding section, if you feel that you have a contribution to make to the cookbook, or a question which you feel ought to be answered in the cookbook, please contact the GSDI Technical Working Group.

Cookbook Overview

The following sections provide an introduction to the content of each chapter. This is provided to help readers decide where to begin their exploration. Some users may already be fluent in geographic information systems but are unfamiliar with the tenets of Spatial Data Infrastructures (SDI). They may wish to start with the next chapter on SDI and GSDI. Others may already have extensive databases that are ready to be published on the World Wide Web. By starting in Chapter Two, they can learn how to catalogue and serve information about their data holdings in standard-based ways.

Chapter 2: Geospatial Data Development: Building data for multiple uses

In Chapter 2, you will learn about the development of standard and non-standard spatial data themes or layers for use in a trans-national or global context. The development of consistent re-usable themes of base cartographic content, known as Framework, Fundamental, Foundation, or Core data is recognized as a common ingredient in the construction of national and global SDIs to provide common data collection schemas.

Chapter 3: Metadata: Describing geospatial data

In Chapter 3, you will learn how geospatial data are documented with metadata, what relevant standards exist, and how to implement them in software. Metadata are a key ingredient in supporting the discovery, evaluation, and application of geographic data beyond the originating organisation or project.

Chapter 4: Geospatial Data Catalogue: Making data discoverable

Geospatial data that are stored for use in local databases can often be used in external applications once they are published. In this chapter, the concepts and implementation of geospatial data catalogues are presented as a means to publish descriptions of your geospatial data holdings in a standard way to permit search across multiple servers.

Geospatial data catalogues are discovery and access systems that use metadata as the target for query on raster, vector, and tabular geospatial information. Indexed and searchable metadata provide a disciplined vocabulary against which intelligent geospatial search can be performed within or among SDI communities.

Chapter 5: Geospatial Data Visualisation: Online Mapping

The primary view of geographic data has historically been through maps. In the context of SDIs, it is increasingly useful to provide mapped or graphical views of geospatial data through online mapping interfaces. This can satisfy many of the needs of novice or browse users of data without requiring download of the full data. 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 at an appropriate scale (*the Fitness for Purpose* concept), 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. This chapter describes current best practice in on-line mapping, and the progress of the OpenGIS Consortium’s Web Mapping Testbed Team in realising the dream of true inter-operability and disseminating a web mapping specification for implementers to adopt and promulgate.

Chapter 6: Geospatial Data Access and Delivery: Open access to data

Once spatial data of interest have been located and evaluated, using the Catalogue and online mapping techniques described in previous chapters, 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. This chapter walks through examples of data access and delivery that are recognized elements in a full-service SDI.

Chapter 7: Other Services

Web mapping services and Catalogue services are described as new, maturing technologies in earlier chapters. Additional services are expected to be identified and selected by multi-organisational projects and will be described and promoted here as they become available.

Chapter 8: Outreach and Capacity Building: Creating a community

The establishment of a Spatial Data Infrastructure at an organisational or national level requires an understanding of the requirements and responsibilities of the members of the community. This chapter discusses, with examples, the elements required for building and sustaining a geospatially-enabled community.

Chapter 9: Case Studies

One of the best ways to articulate the benefits of developing and using a spatial data infrastructure is to highlight the success stories that have emerged at the local, national, regional, and global levels. This chapter provides detailed accounts, or case studies from around the world that put into perspective the value of compatible SDI’s and partnerships in making better decisions regarding the increasingly complex environmental, economic, and social issues that face our communities today.

Chapter 10: Terminology

This chapter will contain the terms used elsewhere in this document with appropriate cross-reference. The abundant use of terms and acronyms in this highly technical field requires such a terminology reference.

Chapter Two: Geospatial Data Development: Building data for multiple uses

Editor: Claude Luzet/MEGRIN; Contributor: Hiroshi Murakami, GSI, Japan, U.S. FGDC

Context and Rationale

In the times of traditional 'mapping', collection, and distribution, of geographic information used to be highly centralised, or controlled by powerful government monopolies. This pattern was established since the beginning of the history of mapping, and lasted for centuries, until very recent times. It was a necessity that had never been challenged due to the heavy costs and technology associated with traditional mapping and to the long time-scales of mapping projects that often extended over several decades. Also maps were not necessarily a consumer product, but were considered part of the national/local assets – data mainly used by the government, for defense, taxes, planning and development.

Thus the governments determined the collection of the information in specific types and formats required for its intended applications. Applications did not vary much across borders, and therefore a similar range of products was developed in many countries. These include:

- Cadastre, cadastral maps (scales from 1:100 to 1:5 000)
- Large scale topographic maps for urban planning and development (scale from 1:500 to 1:10 000)
- National 'base maps' (medium scale, 1:20 000 to 1:100 000)
- Small scale maps (1:100 000 and smaller)

Most, if not all, other mapping products and projects would use these main 'basic maps' as a template, as a common reference, and for building upon this 'basic information' the thematic data and applications that were required. Thus national interoperability was achieved.

Moreover, needs across borders being very comparable, national products across borders were also quite similar, and if edge-matching was not always evident, anyone from country 'A' would be able to read and use a paper map from country 'B with no special effort required. Thus tacit cross-border interoperability also existed.

GIS technology has changed all that, particularly with the development of desktop GIS. Usage and type of applications is now incredibly diverse. GI has become a mass-market product on its own or is found integrated in hard- and software solutions. Nearly anyone can create their own maps, thanks to the use of desktop mapping, GIS, GPS surveying, satellite imagery, scanning and intelligent software. The old monopoly is shaken.

GIS technology is been employed in many different areas and in newer fields of applications, as computer hardware and GIS software applications provide improved capabilities at reduced cost. However, the overall cost of developing geospatial data required to support GIS applications remains relatively high compared with the hardware and software required for GIS.

In addition, GIS users tend to develop their own data sets, even if there are existing geospatial data sets available for them, because:

- they may not know available existing data sets that could be appropriately used for their applications; or access to these data sets is difficult
- they are not used to sharing data sets with other sectors and/or organisations; and
- existing geospatial data sets stored in a certain GIS system may not be easily exported to another system.

These problems arise from the fact that existing geospatial data sets have been poorly documented in a standardised manner. Consequently, there have been duplicate efforts in geospatial data development, which sometimes hinders further dissemination of GIS applications in local, national, regional and global circumstances.

As a result, the new era of GIS is still characterised by:

- many actors involved in data collection and distribution
- a proliferation of GI applications, product types, and formats
- duplication as a consequence of the difficulties to access the existing data, and the highly specific quality of the data collected
- increasing difficulty in the exchange and use of data that came from different organisations

Core-, Reference-, Base-, Fundamental-data, and other similar terms are often used, and generally understood … until one tries to define what concept(s) they cover, or until one tries to define the related specifications.

Most GIS applications employ a limited number of common geospatial data items, including geodetic control points, transportation networks, hydrological networks, contour lines and so forth. These items are common to many GIS applications and provide keys for the integration of other and more specialized thematic information. They represent the content found in most traditional base-maps, or in modern technology and terminology, in most GI databases and products. Does that mean that these items are the ‘core’? What about postal addresses? What about cadastral parcels?

The concepts of ‘core-data’ and of ‘reference-data’ relate to two quite different perspectives. But fortunately they may result in the definition of very similar specifications. Let's start with 'reference'. The primary reference for cartographers is the geodetic and levelling networks that give the surveyors the physical links to a co-ordinate system. Of course, this has recently and dramatically changed with satellite positioning technologies, but the principle remains that the primary reference is what gives access to geodetic coordinates. We are not really concerned with this type of reference here, because it is generally not a part of the Geographic Information that is used in GIS applications, but rather its background. Very often it is even not visible.

If geodesy is the reference for the cartographer and the surveyor, the 'reference' of the GI user is generally more closely related to the real world. It includes concrete themes, such as infrastructure – roads, railways, power-lines, settlements, etc, or physical features – terrain elevation, hydrography, etc. It includes also less tangible features, that have nonetheless a significant role in human life: administrative boundaries, cadastral parcels, gazetteer, postal addresses, etc. All these features are keys that allow one to relate, to 'refer', external information to the real world, through the media of its GI representation. Therefore they may be considered as comprising a reference for the GI user -- the 'reference data'.

A different perspective presides over the conceptual approach of the 'core data'. The core being the heart, the central part, the fundamental part, it may be also considered as being the common denominator of all GI data sets, being so because being used by most applications. We can see that this perspective can bring the specifications of the core very compatible with those deriving from the concept of the 'reference data'. Therefore, let's not loose ourselves in academic debates, and let's keep here a simple practical view and terminology.

'Core data' when used here, will mean "a set of Geographic Information that is necessary for optimal use of most GIS applications, i.e. that is a sufficient reference for most geo-located data." The relevance of this definition can of course be questioned, and will need to be improved. Let's adopt it only for the sake of understanding the following chapters. One obvious necessary accommodation to the above definition, is that the specifications might be scale-dependent. Core, then, may refer to the fewest number of features and characteristics required to represent a given data theme.

We have seen before that the GIS revolution has resulted in a democratisation of GI, but also in a key problem that is the non-interoperability of the GI produced with the new technologies. We propose that the concept of the 'core data' is one instrumentality to help improving interoperability, thus increasing GI usability and reducing expenses resulting from the current duplications.

Interoperability complications exist at different levels, and they can be found in four main types:

- cross-border : edge matching between different data sets
- cross-sector : data sets created for different sector-based applications
- cross-type : e.g. raster- vs. vector-data

- overlap : same features coming from different sources and process

Resolving the related issues will need a mix of three ingredients -- the technology, the adoption of a common concept of 'core data', and of course the political support that will help resourcing the necessary key implementations.

The concept of the core aims at sharing the core data sets between users in order to facilitate the development of GIS. Each data item may be provided by a different data provider. Such data providers produce data through their daily businesses including road management, urban planning, land management, tax collection, and so forth. Although there may be many data providers, the data sets they provide must be integrated to develop core data sets. Once these core data sets are shared between data users, each user does not have to develop the core data by oneself, and can avoid duplicated efforts of core data development. Consequently, by sharing the cost of developing the core data, data development cost can be minimised and shared between users.

Much more than at the time of data set creation, the benefits of the 'core data' concept will be revealed when updating. Since these core data sets are developed by those who produce the data through their daily businesses, they are updated most frequently. Therefore, the users are assured of using up-to-date core data sets. In addition, these data producers develop most detailed geospatial data with high quality based on their business requirements. Another benefit of using core data sets lies in the fact that these commonly used core data sets enable the users to easily share other geospatial data with other users.

Achieving Benefits

In order to achieve those benefits described in the previous section, those data producers who develop and maintain geospatial data sets through their daily businesses are to distribute their data to the public. Once distributed, GIS users can collect and integrate them in their own GIS applications. Such data sets would provide GIS users with the most up-to-date and highest quality data sets publicly available. Hence the users have to spend only a minimum amount of cost for the core data in their GIS applications.

Global Map is one illustration of 'core' data sets conceived in a global or at least multi-national environment. The Japanese Geographical Survey Institute took an initiative in 1992 to develop a suite of global geospatial data (Global Map) to cope with the global environmental problems. The goal is to involve national mapping organisations to collaboratively develop global geospatial data sets. By incorporating national mapping organisations of the world, the collected information would be most up-to-date and assured of being free of national security issues. The Global Map could be considered as an initial implementation of the concept of a suite of 'core data' for GSDI in concert with similar framework data sets at regional and national levels.

It is important to recognize that Core data, as represented by Global Map and other national initiatives, do not comprise the only data available within a national or global SDI. SDI capabilities enable the documentation and service of all types of geospatial data, such as local scientific or engineering projects, regional or global remote sensing activities, and environmental monitoring. Although SDIs as infrastructure enables access to all these types of information, special consideration is given in this chapter to document issues associated with data of high re-use potential that may be served by SDIs at local, national, or global levels as traditional base map themes.

Organisational Approach

At the national level, common spatial data are often defined through community and/or national agreements on content, known as "framework" or "fundamental" data in various national SDIs. In the Australian Spatial Data Infrastructure (ASDI), Fundamental describes a dataset for which several government agencies, regional groups and/or industry groups require a comparable national coverage in order to achieve their corporate objectives and responsibilities. In other words, fundamental data are a subset of framework data. Similar concepts exist in other countries with similar terms, and most identify general themes of interest as "framework" information, for they provide a framework of base, common-use geospatial information onto which thematic information can be portrayed. An organisation interested in implementing spatial data that will be compatible with local, regional, national, and global data sets, must identify, and potentially reconcile different framework designations across their geographic area of interest.

The framework is a collaborative effort to create a common source of basic geographic data. It provides the most common data themes geographic data users need, as well as an environment to support the development and use of these data. The framework's key aspects are:

- specific layers of digital geographic data with content specifications
- procedures, technology, and guidelines that provide for integration, sharing, and use of these data; and
- institutional relationships and business practices that encourage the maintenance and use of data.

The framework represents a foundation on which organisations can build by adding their own detail and compiling other data sets. Existing data content may be enhanced, adjusted, or even simplified to match a national or global framework specification. This is helpful for the purpose of exchange.

Framework Leverages the Development of Needed Data

Thousands of organisations spend billions of dollars each year producing and using geographic data. Yet, they still do not have the information they need to solve critical problems. There are several aspects to this problem:

- Most organisations need more data than they can afford. Frequently, large amounts of money are spent on basic geographic data, leaving little for applications data and development.
- Some organisations cannot afford to collect base information at all. Organisations often need data outside their jurisdictions or operational areas. They do not collect these data themselves, but other organisations do.
- Data collected by different organisations are often incompatible. The data may cover the same geographic area but use different geographic bases and standards. Information needed to solve cross-jurisdictional problems is often unavailable.

Many of the resources organisations spend on geographic information systems (GIS) go toward duplicating other organisations' data collection efforts. The same geographic data themes for an area are collected again and again, at great expense. Most organisations cannot afford to continue to operate this way.

Framework initiatives will greatly improve this situation by leveraging individual geographic data efforts so data can be exchanged at reasonable cost by government, commercial, and non-governmental contributors. It provides basic geographic data in a common encoding and makes them discoverable through a catalogue (See Chapter 4) in which anyone can participate. Using Web mapping and advanced, distributed GIS technology in the future, users can perform visual cross-jurisdictional and cross-organisational analyses and operations, and organisations can funnel their resources into applications, rather than duplicating data production efforts.

There are many situations in which the framework will help users. A regional transportation planning project can use base data supplied by the localities it spans. Government agencies can respond quickly to a natural disaster by combining data. A jurisdiction can use watershed data from beyond its boundaries to plan its water resources. Organisations can better track the ownership of publicly held lands by working with parcel data.

Geographic data users from many disciplines have a recurring need for a few themes of basic data. While these layers may vary from place to place, some common themes include: geodetic control, orthoimagery, elevation, transportation, official geographic names (gazetteer), hydrography, governmental units, and cadastral information. Many organisations produce and use such data every day. The framework provides basic content for these data themes, and by defining a common schema, it can also provide a common means of information exchange and value-adding.

By attaching their own geographic data — which can cover innumerable subjects and themes — to the common data in the framework, users can build their applications more easily and at less cost. The common data themes provide basic data that can be used in applications, a base to which users can add or attach geographic details and attributes, reference source for accurately registering and compiling participants' own data sets, and a reference map for displaying the locations and the results of an analysis of other data.

National and global frameworks are a growing data resource to which geographic data producers can contribute. It will continually evolve and improve. In practice, the content model of many framework layers may be simple enough that, as a collection target, at certain scales, it could be made available at virtually no cost. Content providers exist already in the United States to take and extend free government data with valuable additional attributes of value, e.g. marketing and demographic information. The core information itself may be given away for free, but extended information that are anchored to the geometry may have high current value that declines over time, and may re-enter the public domain after its proprietary nature expire. Thus commercial providers of information benefit through anchoring to a common framework system and cross-referencing with other attributes held by other organisations; consumers benefit in acquiring the framework core geometry, feature definitions, and base attributes as a by-product of the more advanced data set.

Who are the actors in framework data development?

- users and producers of detailed data, such as utilities
- users of small-scale, limited geographic data, such as street networks, statistical areas, and administrative units;
- data producers who create detailed data as a product or a service;
- data producers who create low-resolution, small-scale, limited themes for large areas;
- product providers who offer software, hardware, and related systems; and
- service providers who offer system development, database development, operations support, and consulting services.

Non-profit and educational institutions also create and use a variety of geographic data and provide GIS-related services. They cover the full spectrum of data content, resolution, and geographic coverage. Depending on the organisation's activities, data use may range from high-resolution data over small areas, as in facility management, to low-resolution data over wide areas, as in regional or national environmental studies.

Organisations build national and regional framework efforts by coordinating their data collection and development activities based on intersecting interests within a community. The bounds of this community, however, given the diversity of types of organisations and individuals involved, needs to be non-exclusive and open to innovative contributions, exchanges, and partnerships. The framework should be developed by the entire community, with organisations from all areas playing roles. For some, the framework will supply the data they need to build applications. Others will contribute data, and some may provide services to maintain and distribute data. Some organisations will play several roles in framework development, operation, and use. The framework will take many years to develop fully, but useful components are being developed continuously.

Implementation Approach

The ISO TC 211 Geomatics standardisation activity is working on two related areas of endeavour that will greatly assist in the global specification of content models and feature models for framework and non-framework data. These include ISO 19109 - Rules for application schema, and 19110 - Feature catalogueuing methodology. In the networked world, the ability for software to interact with geographic information outside an organisation is virtually non-existent except where public agreements exist for data structures (also known as a content model or schema) and the features being mapped. The ISO standards mentioned above provide a basis for the description of these packages of information that would enable access to a distributed network of framework data services. Coupled with catalogue for discovery (See Chapter 4) populated with metadata (See Chapter 3), the ingredients are coming together for a configurable deployed architecture.

The scope of ISO 19109 is defined as "… the rules for defining an application schema, including the principles for classification of geographic objects and their relationships to an application schema." Using the Unified Modeling Language, software applications that provide access to geospatial data, such as framework, would be defined in a consistent way so as to improve sharing of data between applications and even allow for real-time interaction between applications.

Before one can allow software to reliably access mapped features stored in remote data systems, there must first be a common understanding about the nature and composition of the objects being managed. ISO 19109 includes guidance principles for classifying geographic objects. The usefulness of any information is reduced when the meaning is unclear, especially and commonly across different application domains. If different classifications are defined using a consistent set of rules, that ability to map one classification to another and retain the meaning will be greatly increased. This is also known as the semantic translation of one representation of an object in one system, for example a road or river segment, to that in another.

These rules will be used by geographic information users when classifying geographic objects within their applications and when interpreting geographic data from other applications. The rules and principles could also be used by geographic information system and software developers to design tools for the creation and maintenance of classification schemes.

Very closely related to the schema definition of ISO 19109 is the standard proposing a feature catalogueuing methodology, ISO 19110. It is intended to define the approach and structures used for an information provider to store the identity, meaning, representation, and relationships of concepts or things in the real world as they are managed in online systems. A feature catalogue, then, acts as a dictionary for feature types or classes that can be used in software. The definition of a single international, multilingual catalogue would have tremendous value. Whether this catalogue was used in all applications or only used as a neutral form when moving data from one application to another, it could simplify the problem of mapping the catalogue of one application to the catalogue of another. However, the feasibility of such a task is in question and will be investigated as a part of this work item in the TC 211 work group. The catalogueing task will use the input from the Rules for Application Schema work item and cannot be completed before that item is completed.

Publishing an application schema with a feature catalogue for a given data set of common interest can provide the basis for framework data definitions of use to global, regional, national, and local data. Done carefully, schemas and feature catalogueues could be similarly constructed for existing framework-like data in order to enable discussion among participants, and transformation of content into conforming framework data sets.

A project to develop framework specifications in Switzerland, known as InterLIS, has had marked success with this approach. Common definitions of data layers exist as target specifications that are matched to various degrees by participant organisations. As a result, software that is designed to interact with the InterLIS application model will work against data sets from different sources and organisations. The application framework is designed to be a scalable one to allow the participation of minimal data sets with lesser application functionality and more complex data sets with maximal application functionality.

Common Identities of Real World Objects

In many framework implementations, there will not be necessarily one authoritative geometric representation of a feature in the real world. Several national systems have proposed the use of a common or permanent feature identifier to be associated with the object in the real world so that different representations and attributes of that object on maps can be cross-referenced. Having well-known identities of features established with a coding system within a community greatly assists in the association of attribute information to real-world objects where such attributes may not reside in a GIS or spatially-enabled data base. Also, multiple representations of real world objects may be linked to the identity code, to provide views of an object that is changed over time or that has different degrees of spatial resolution at different scales of data collection or representation. This becomes a logical model for organizing related geospatial information.

The management of a common or "permanent" feature identity needs to be undertaken within the community with permission granted to certain participant organisations to create or adjudicate these identities. In Canada, there is an effort to create a data alignment layer of well-known features or intersections of features to help vertically integrate spatial data from different sources. These features and intersections will have published identifiers, some sense of positional accuracy, and source information. In the United States, the National Hydrography Data set includes a permanent feature identifier for segments of river and water bodies between points of confluence. In other national, regional, and global settings, agreement on management and assignment of feature identifiers -- building upon a sound feature catalogueuing approach -- will be essential in building up compatible framework data across political boundaries.

Candidate National Framework Categories

A variable number of data layers may be considered to be common-use and of national or trans-national importance as "framework" data. Framework layers commonly nominated in national context include:

- cadastral information
- geodetic control
- geographic feature names
- orthoimagery
- elevation
- transportation
- hydrography
- governmental units

It is likely for this list to grow as custodians of data identify and promote their data as necessary to increasingly advanced applications and user environments.

Candidate Global Data Categories

The Global Mapping concept was articulated by the Ministry of Construction of Japan as a response to the United Nations Conference on Environment and Development held in Brazil in 1992. Agenda 21 is an action program drawn up by the conference, and it clearly makes the case that global baseline spatial data is important to society's interaction with the environment. The Global Mapping Project, also known as Global Map, is addressing the compilation of suitable spatial data products from existing international and national sources. This provide a public set of reference data at trans-national to global scales to assist decision-makers and society in depicting global environmental concerns.

Progress is being made in selecting and enhancing these general purpose spatial data layers originally based on VMAP Level 0 (also known as Digital Chart of the World) for vector themes, Global Land Cover Characteristics Database from the U.S. Geological Survey (USGS) for land cover, land use and vegetation, and the 30-second GTOPO30 product also hosted by the USGS. Global Map Version 1.0 specifications for data organisation were adopted at the International Steering Committee for Global Mapping (ISCGM) meeting held in conjunction with the Third GSDI Conference in Canberra, Australia in November 1998. As of February 2000, 74 countries are participating in the collection or aggregation of large-scale map products to update and package the above data sources.

Recommendations

The development of common data specifications is an arduous task to undertake by oneself or by a single organisation. For the development of the GSDI the following recommendations are made:

- ***The Cookbook authors recommend that interested parties participate in or be aware of existing framework initiatives at the sub-national, national, and international scale.***

Data appropriate to a given type of geospatial analysis will require information at a range of resolutions and degrees of detail.

The Cookbook authors recommend that Global Map specification be adopted for trans-national applications requiring land cover/use, vegetation, transportation, hydrography, administrative boundaries, populated places, and elevation data.

The global map content specification defines a simple content model with a small number of feature types and attributes suitable for the construction of base cartography at regional scales. Evaluate the level of detail with respect to a given GIS or mapping application. It may require extension to suit your base requirements.

- ***The Cookbook authors recommend that Core and non-Core data be modeled and shared in the designs of national SDIs using emerging ISO standards.***

The Cookbook authors suggest prototyping and review of pending ISO 19109 and 19110 standards on application schema and feature catalogueues for use in GIS applications for all types of data (not just Core).

The ISO work in progress formalises the description of features and feature collections for individual applications that can facilitate the proper access and transformation of geospatial data held in online systems in near real time. This extends the capabilities of the individual in working with dynamic information held in distributed locations, as will be discussed in Chapter 6 in greater detail. National and global framework data, as well as non-framework data will be made more accessible and semantically correct through such technologies.

References and Linkages

Australian Spatial Data Infrastructure Fundamental Data
http://www.auslig.gov.au/asdi/ffdata.htm

Framework Home Page, U.S. Federal Geographic Data Committee
http://www.fgdc.gov/framework/framework.html

Global Mapping Specifications - Version 1.0, 20 November 1998
http://www.auslig.gov.au/mapping/global_m/specv1_0.htm

Interlis Project Home Page, Switzerland
http://www.gis.ethz.ch/interlis/index_e.html

Chapter Three: Metadata -- Describing geospatial data

Editor: Mark Taylor, United Kingdom, NGDF
This document has been developed from input by FGDC, EUROGI, ANZLIC and NGDF and is predominantly based on the various sources cited at the end of the chapter.

Introduction

We often hear the phrase "information is power," but with increasing amounts of data being created and stored (but often not well organised) there is a real need to document the data for future use - to be as accessible as possible to as wide a "public" as possible. Data, plus the context for its use (documentation, metadata) become information. Data without context are not as valuable as documented data. There are significant benefits to such asset management:

- Metadata helps organise and maintain an organisation's investment in data and provides information about an organisation's data holdings in catalogueue form
- Coordinated metadata development avoids duplication of effort by ensuring the organisation is aware of the existence of data sets
- Users can locate all available geospatial and associated data relevant to an area of interest
- Collection of metadata builds upon and enhances the data management procedures of the geospatial community
- Reporting of descriptive metadata promotes the availability of geospatial data beyond the traditional geospatial community
- Data providers are able to advertise and promote the availability of their data and potentially link to on line services (e.g. text reports, images, web mapping and e-commerce) that relate to their specific data sets

A number of studies have established that although the value of geospatial data is recognised by both government and society, the effective use of geospatial data is inhibited by poor knowledge of the existence of data, poorly documented information about the data sets, and data inconsistencies. Once created, geospatial data can be used by multiple software systems for different purposes. Given the dynamic nature of geospatial data in a networked environment, metadata is therefore an essential requirement for locating and evaluating available data. Metadata can help the concerned citizen, the city planner, the graduate student in geography, or the forest manager find and use geospatial data, but they also benefit the primary creator of the data by maintaining the value of the data and assuring their continued use over a span of years. Over thirty years ago, humans landed on the Moon. Data from that era are still being used today, and it is reasonable to assume that today's geospatial data could still be used in the year 2020 and beyond to study climate change, ecosystems, and other natural processes. Metadata standards will increase the value of such data by facilitating data sharing through time and space. So when a manager launches a new project, investing a small amount of time and resources at the beginning may pay dividends in the future.

Context and Rationale

The word metadata shares the same Greek root as the word metamorphosis. "Meta-" means change and metadata, or "data about data" describe the origins of and track the changes to data. Metadata is the term used to describe the summary information or characteristics of a set of data. This very general definition includes an almost limitless spectrum of possibilities ranging from human-generated textual description of a resource to machine-generated data that may be useful to software applications.

The term *metadata* has become widely used over the past 15 years, and has become particularly common with the popularity of the World Wide Web. But the underlying concepts have been in use for as long as collections of information have been organised. Library catalogueues represent an established variety of metadata that has served for decades as collection management and resource discovery tools. The concept of metadata is also familiar to most people who deal with spatial issues. A map legend is one representation of metadata, containing information about the publisher of the map, the publication date, the type of map, a description of the map, spatial references, the map's scale and its accuracy, among other things. Metadata are also these types of descriptive information applied to a digital geospatial file. They're a common set of terms and definitions to use when documenting and using geospatial data. Most digital geospatial files now have some associated metadata. In the area of geospatial information or information with a geographic component this normally means the What, Who, Where, Why, When and How of the data. The only major difference that therefore exists from the many other metadata sets being collected for libraries, academia, professions and elsewhere is the emphasis on the spatial component - or the where element.

The Benefits of Metadata

Metadata helps people who use geospatial data find the data they need and determine how best to use it. Metadata benefit the data-producing organisation as well. As personnel change in an organisation, undocumented data may lose their value. Later workers may have little understanding of the contents and uses for a digital database and may find they can't trust results generated from these data. Lack of knowledge about other organisations' data can lead to duplication of effort. It may seem burdensome to add the cost of generating metadata to the cost of data collection, but in the long run the value of the data is dependent on its documentation.

Metadata is one of those terms that is conveniently ignored or avoided. However there is an increasing recognition of the benefits and requirement for metadata for our data as we continue to increase the use of digital data. Whereas cartographers rigidly provided metadata within a paper map's legend, the evolution of computers and GIS has seen a decline in this practice. As organisations start to recognize the value of this ancillary information, they often begin to look at incorporating metadata collection within the data management process.

Organisational Approach

Levels of Metadata

There are different levels that metadata may be used for:

- Discovery metadata - What data sets hold the sort of data I am interested in? This enable organisations to know and publicise what data holdings they have.
- Exploration metadata - Do the identified data sets contain sufficient information to enable a sensible analysis to be made for my purposes? This is documentation to be provided wit the data to ensure that others use the data correctly and wisely.
- Exploitation metadata – What is the process of obtaining and using the data that are required? This helps end users and provider organisations to effectively store, reuse, maintain and archive their data holdings.

Each of these purposes, while complementary, requires different levels of information. As such organisations should look at their overall needs and requirements before developing their metadata systems. The important aspect is for agencies to establish their business requirements first, the content specifications second and the technology and implementation methods third.

This is not to say that these levels of metadata are unique. There is a high degree of reuse of the metadata for each level and an organisation will design its metadata schema and implementation based on its business needs to accommodate these three requirements.

Discovery Metadata is the minimum amount of information that needs to be provided to convey to the inquirer the nature and content of the data resource. This falls into broad categories to answer the "what, why when who, where and how" questions about geospatial data.

What - title and description of the data set.
Why - abstract detailing reasons for the data collection and its uses.
When - when the data set was created and the update cycles if any.
Who – originator, data supplier, and possibly intended audience.
Where - the geographical extent based on latitude / longitude, co-ordinates, geographical names or administrative areas.

How – how it was built and how to access the data.

The broad categories are only few in number to reduce the effort required to collect the information whilst still conforming to the requirement to convey to the inquirer the nature and content of the data resource.

Online systems for handling metadata need to rely on their (*metadata* is plural, like *data*) being predictable in both form and content. The level of metadata detail that will be documented is dependent on the type of data held and the methods that it is being accessed and used. Different types of data (e.g. vector, raster, textual, imagery, thematic, boundary, polygon, attribute, point, etc.) will require different levels and forms of metadata to be collected. However there is still a high degree of compatibility between most of the metadata elements required.

Similarly, organisations will manage their data in mission-defined ways. Some organisations manage information as a data set, tiles of data sets, series of data sets, or manage the information down to the feature level. Again there is still a high level of compatibility between the levels of metadata required, particularly as the data is cascaded from the feature level to the data set or data series level.

Thus, not only can metadata content vary according to purpose; it can also vary according to scope of the data being defined. Discovery metadata usually, but not exclusively, relates to collections of data resources or data set series that have similar characteristics but relate to different geographic extents or times. A map series is the commonest example but it can equally be applied to statistical surveys. More detailed metadata may be applied to a collection or series but may apply to an individual data set (e.g. one map tile). Transfer metadata applies exclusively to that transfer.

Exploration metadata provides sufficient information enable an inquirer to ascertain that data fit for a given purpose exists, to evaluate its properties, and to reference some point of contact for more information. Thus, after discovery, more detail is needed about individual data sets, and more comprehensive and more specific metadata is required. If the data are transferred as a single data set then quite specific and detailed metadata is needed possibly down to the feature, object or record level. Exploration metadata include those properties required to allow the prospective end user know whether the data will meet general requirements of a given problem.

Exploitation metadata include those properties required to access, transfer, load, interpret, and apply the data in the end application where it is exploited. This class of metadata often includes the details of a data dictionary, the data organisation or schema, projection and geometric characteristics, and other parameters that are useful to human and machine in the proper use of the geospatial data.

These roles form a continuum in which a user cascades through a pyramid of choices to determine what data are available, to evaluate the fitness of the data for use, to access the data, and to transfer and process the data. The exact order in which data elements are evaluated, and the relative importance of data elements, will not be the same for all users.

Linkages between geospatial data and metadata

Until recently, metadata have been created or derived with little or no automation. In fact, it is only with the recent development of metadata standards, and the development of metadata software based on these standards, has the consistent management of metadata been given any consideration by those collecting geospatial data. With an increased focus of incorporating geospatial data into corporate information systems, the development of an international standard for metadata, and the OpenGIS catalogueue service specifications, new versions of commercial GIS software are now facilitating a close linkage between geospatial data and metadata.

Regardless of style of metadata, there is nominally one collection of properties or metadata associated with a given data set or feature collection. The 1:1 rule expresses the notion that a discrete resource should have a discrete metadata record. Although it seems simple enough, it isn't always so neat because resources are often not so discrete. For example, should each photograph in an article have its own record? How do you manage collections of articles? Can the collection be thought of as a resource? What about multi-media objects? Thus, one of the first tasks in metadata management is the identification of the data product or entity to be documented.

Metadata may exist at the collection level (e.g. satellite series), at a data product level (an image mosaic), at a data unit level (a vector data set), a group of features of a given type (certain roads), or even at a specific feature instance (a single road). Regardless of the level of abstraction, these associations of metadata to data objects should be maintained.

In practice, most metadata are currently collected at the data set level, and a metadata entry in a catalogueue refers the user to its location for access. Increasingly sophisticated providers of geospatial data are including metadata at other levels of detail so as to preserve information richness. Metadata standards such as ISO 19115 allow different levels of metadata abstraction, and catalogueue services will also need to accommodate this richness without confusing the user in its complexity.

Metadata Standards

Why use Standards?

Ideally, metadata structures and definitions should be referenced to a standard. One benefit of standards is that they have been developed through a consultative process (with other "experts") and provide a basis from which to develop national or discipline-oriented profiles. As standards become adopted within the wider community, software programs will be developed to assist the industry in implementing the standard. The consistency in metadata content and style is recommended to ensure that comparisons can be made quickly by data users as to the suitability of data from different sources. This means for example when comparing metadata about property or hazardous waste there is an indication of the dates to which the information refers or if comparing metadata about different map sources the relevant scales are shown. Without standardization, meaningful comparisons are more difficult to derive without reading and learning many metadata management styles.

Predictability is also encouraged through conformance to standards. However the problem has been that there are a number of "standards" in use or development. Detailed metadata standards that provide for an exhaustive definition of all aspects of various types of geospatial data are currently under preparation by a number of bodies, as are profiles of these standards as reference models to be adopted for international use.

Geospatial Metadata Standards

Considerable debate across the world centres on metadata and those characteristics that should be chosen to best describe the data set. There are discussion groups, seminars and conferences and quantities of paper generated in the debate about the subject. Standards have been generated by a number of organisations all designed to ensure that a degree of consistency exists within a given application community.

Three main metadata standards exist or are in development that are of broad international scope and usage and provide detail for all levels of metadata mentioned earlier:

The **Content Standard for Digital Geospatial Metadata**, U.S. 1994, revised 1998
http://www.fgdc.gov/

In the USA the Federal Geographic Data Committee (FGDC) approved their Content Standard for Digital Geospatial Metadata in 1994. This is a national spatial metadata standard developed to support the development of the National Spatial Data Infrastructure. The standard has also been adopted and implemented in the United States, Canada, and the United Kingdom through the National Geographic Data Framework (NGDF). It is also in use by the South African Spatial Data Discovery Facility, the Inter-American Geospatial Data Network of twelve Latin American countries, and elsewhere in Asia.

A **CEN Pre-standard** adopted in 1998
http://forum.afnor.fr/afnor/WORK/AFNOR/GPN2/Z13C/indexen.htm

In 1992 the Comité Européen de Normalisation (CEN) created technical committee 287 with responsibility for geographic information standards. A family of European

Pre-standards have now been adopted including 'ENV (Euro-Norme Voluntaire) 12657 Geographic information - Data description - Metadata'.

ISO TC 211 Standard (19115-Draft International Standard)

An ISO standard is now at approval stage and expected to be ratified in late 2001 (http://www.statkart.no/isotc211/welcome.html). In 1994 the International Standards Organisation created technical committee 211 (ISO/TC 211) with responsibility for Geoinformation/Geomatics. They are preparing a family of standards; this process involves a working group, a committee draft, a draft international standard and finally the international standard. ISO have now released the committee draft of 'ISO 19115 - GI – Metadata. It is hoped that all the existing standards will converge through the ISO initiative. Indeed, most of the existing standards already have a great deal in common and a robust international discussion has ensured that the ISO standard has accommodated most of the various international requirements. The ISO standard has equally benefited from the experiences of the various national bodies and their implementations of the respective metadata standards assisted by metadata software.

Metadata also forms an important part of the OpenGIS Abstract Specification. The OpenGIS Consortium (OGC) *http://www.opengis.org* is an international membership organisation engaged in a co-operative effort to create open computing specifications in the area of geoprocessing. As part of its draft 'OpenGIS Abstract Specification' OGC has a topic on recording metadata for spatial data. OGC are working closely with FGDC and ISO/TC 211 to develop formal, global spatial metadata standards. At their plenary meeting in Vienna, Austria in March 1999, ISO/TC 211 welcomed the satisfactory completion of the co-operative agreement between the OpenGIS Consortium and ISO/TC 211 and endorsed the terms of reference for an ISO/TC 211 / OGC co-ordination group.

These standards have had different ideas about what characteristics should be included. To derive all these elements the data provider requires spending considerable time and resources collecting this information and for the data user this detail might be greater than required for an initial investigation. In many situations therefore different levels of metadata need to be defined with the ability to "drill down" into increasing levels of detail. Metadata should therefore vary according to purpose.

A number of national and regional initiatives have also developed metadata standards. These include initiatives managed by The Australian and New Zealand Land Information Council (ANZLIC) and two European Commission financed projects (LaCLEF and ESMI). These initiatives have taken similar approaches in promoting a limited set of metadata (described as "Core Metadata" or "Discovery Metadata" that organisations should use, as a minimum, to improve the knowledge, awareness and accessibility of the available geospatial data resources.

Each of the initiatives is promoting the standards and use of discovery metadata as a foundation of their respective metadata directory initiatives. This discovery metadata provides sufficient information to enable an inquirer to ascertain that existence of data fit for purpose exists and to reference some point of contact for more information. If, after discovery, more detail is needed about individual data sets then more comprehensive and more specific metadata is required. It is possible that organisations may wish to develop metadata at different but complementary levels - at one level discovery metadata for external use and for in-house / internal use more detailed metadata. And to avoid duplication of effort those elements common to both are flagged. These guidelines have been developed with recognition of the importance of more extensive metadata required for data management and each of the organisations is promoting the adoption of ISO Metadata Standard.

General Metadata Standards

Other standards exist in the broader topic of metadata that do not specifically apply to geospatial information. These conventions are listed here for informational purposes. They may be useful references for linking or integrating non-geospatial resources into a geospatial framework.

The Dublin Core is a metadata element set intended to facilitate discovery of electronic resources. Originally conceived for author-generated description of Web resources, it has attracted the attention of formal resource description communities such as museums, libraries, government agencies, and commercial organisations.

The Dublin Core Workshop Series has gathered participants from the library world, the networking and digital library research communities, and a variety of content specialists in a series of invitational workshops. The building of an interdisciplinary, international consensus around a core element set is the central feature of the Dublin Core. The progress represents the emergent wisdom and collective experience of many stakeholders in the resource description arena. Dublin Core metadata is specifically intended to support general-purpose *resource discovery.* The elements represent one community's concepts of core elements that are likely to be useful to support resource discovery. Unfortunately, the formal use of the Dublin Core metadata model does has not always recognized the inclusion of *qualified* elements such as “Coverage.” This metadata element thus may contain text that represents a date or time, a description of a place name or time period, or coordinates, without a means to declare what type of content is present in the text element. As such, the Dublin Core unqualified elements are inadequate for even basic geospatial resource description and discovery, though they may be applied to web and library resources with a loose geospatial definition. Qualified Dublin Core elements can be derived from more detailed metadata models (such as ISO 19115) and can support discovery of lightly documented ancillary information such as books, reports, and other Web objects of potential interest to geospatial investigations.

The Spatial Data Transfer Standard (SDTS) and the Vector Product Format (VPF) Digital Exchange Standards (DIGEST) were developed to allow the encoding of digital spatial data sets for transfer between spatial data software. Both of these standards support the inclusion of metadata elements in an exchange, but remarkably have not until recently considered support for standardised the encoding of relevant geospatial metadata standards in their export or archival formats.

While other general-purpose metadata standards exist, it is recommended that a comprehensive geospatial metadata standard should be used to document geospatial data. It is easier to produce simplified metadata from a more robust collection of metadata, but it is impossible to do the opposite. Eventually, the integration of data content and exchange standards will converge with those in metadata content and exchange so that spatial data encoding efforts will provide a comprehensive solution for archive and documentation.

Implementation Approach

Who should create metadata?

Data managers tend to be either technically literate scientists or scientifically literate computer specialists. Creating correct metadata is like library catalogueuing, except the creator needs to know more of the scientific information behind the data in order to properly document them. Don't assume that every professional needs to be able to create proper metadata. They may complain that it is too hard and they may not recognise the benefits. In this case, ensure that there is good communication between the metadata producer and the data producer; the former may have to ask questions of the latter to collaboratively develop adequate documentation.

The form for maintaining metadata will depend on a number of factors:

- the size of the data holdings,
- the size of an organisation and
- the patterns of data management within an organisation

If the metadata holdings are fairly modest, then it has been the convention to store the metadata in discrete documents by using any available software (e.g. word-processor, spreadsheet, and simple database). Historically, organisations have built up folders of single documents that may be in either paper or digital formats. Many organisations will start to investigate the use of more complex systems as they realise the benefit of the metadata, and as they gain greater data holdings and start to provide broader access to the data.

Indeed many organisations will start with a basic audit of their data holdings that will alert them of the vast wealth of data that they possess and where it is being used, replicated or improved across the organisation. As the data holdings become larger and the access to the data becomes distributed, then organisations would look at more advanced methods for maintaining metadata of their data holdings. These advanced tools may consist of commercial or self-developed forms based systems that may also form part of the operational GI systems to extract aspects of the metadata automatically from the data itself.

How does one deal with people who complain that it's too hard? The solution in most cases is to redesign the workflow rather than to develop new tools or training. People often assume that data producers must generate their own metadata. Certainly they should provide informal, unstructured documentation, but they may not need to go through the rigors of fully structured formal metadata. For scientists or GIS specialists who produce one or two data sets per year it may not be worth their time to fully learn a complex metadata standard. Instead, they might be asked to fill out a less-complicated form or template that will be rendered in the proper format by a data manager or catalogueuer who is familiar (not necessarily expert) with the subject and well-versed in the metadata standard. If twenty or thirty scientists are passing data to the data manager in a year, it *is* worth the data manager's time to learn the complex metadata standard. With good communication, this strategy complements the existing combination of software tools and training.

The first data set documented is always the worst. The other aspect to "It's too hard" is that documenting a data set fully requires a (sometimes) uncomfortably close look at the data and brings home the realisation of how little is really known about its processing history.

"Insufficient time" to document data sets is also a common complaint. This is a situation in which managers who appreciate the value of GIS data sets can set priorities to protect their data investment by allocating time to document it. Spending one or two days documenting a data set that may have taken months or years to develop at thousands of dollars in cost hardly seems like an excessive amount of time.

These 'pain' and 'time' concerns have some legitimacy, especially for agencies that may have hundreds of legacy data sets which could be documented, but for which the time spent documenting them takes away from current projects. At this point in time, it seems much more useful to have a lot of 'shortcut' metadata rather a small amount of full-blown metadata. So what recommendations can be made to these agencies with regard to a sort of 'minimum metadata' or means to reduce the documentation load?

In some operations, small amounts of metadata, or “notes” are collected sporadically during the data processing flow. These hints can then be assembled more readily later into a clear statement of the history and processing of the dataset. This can present a less daunting task at the end of a project as most of the details are already documented, a little at a time. Increasingly, GIS and image processing software are capable of collecting and reporting quantitative metadata that can be filled-in for the user rather than expecting human input. These procedures can amount to significant savings in overall time and effort over a single manual metadata preparation process at the conclusion of a project.

Don't invent your own standard. Select a supported international standard wherever possible. Try to stay within its constructs. Subtle changes from an international standard such as collapse of compound elements may be costly in the long run - you won't be able to use standard metadata tools and your metadata may not be directly exchangeable or paresable by software.

Don't confuse the metadata presentation (view) with the metadata itself. There is a temptation to lump form and content into the same bin (e.g. "What I see in my database is what I print"). However, the ability to differentiate the contents of the metadatabase (the columns or fields) from its presentation (writing formatted reports) is now commonplace in desktop database software packages. This allows users to consider more flexibly *how* to present *what* information.

There are typically three forms of metadata that should be recognized and supported in systems: the implementation form (within a database or software system), the export or encoding format (a machine-readable form designed for transfer of metadata between computers), and the presentation form (a format suitable to viewing by humans). By recognizing the connections between these dispositions of metadata, one can build systems that support mission requirements, standard encoding for exchange, and permit many “report” views of the metadata to satisfy the needs and experience of different user constituencies.

The Extensible Markup Language (XML) provides two solutions to this metadata problem. First, it includes a capable markup language with structural rules enforced through a control file to validate document structure. Second, through a companion standard (XML Style Language, or XSL), an XML document may be used along with a style sheet to produce standardised presentations of content, allowing the user to shuffle field order, change tag names, or show only certain fields of information. Used together XML and style sheets allow for a structured exchange format and for flexible presentation. Thus, a metadata entry can be rendered in many ways from the same, single structured encoding.

XML is a widely accepted encoding methodology with international software support. It is supported by a lot of software, both free and commercial. However, the metadata-producing community doesn't have much experience using it to solve problems yet. Through reference implementations of software and experimentation, local Spatial Data Infrastructures can share their successes and failures in applying this new technology to fullest community benefit.

Consider data granularity. Can you document many of your data sets (or tiles) under an umbrella parent? Prioritise your data. Begin by documenting those data sets that have current or anticipated future use, data sets that form the framework upon which others are based, and data sets that represent your organisation's largest commitment in terms of effort or cost.

Document at a level that preserves the value of the data within your organisation. Consider how much you would like to know about your data sets if one of your senior GIS operators left suddenly in favour of a primitive lifestyle on a tropical island.

How do I create metadata?

First, one should understand both the data you are trying to describe and the standard itself. Then one must decide how you to encode the information. Historically, one creates a single text file for each metadata record; that is, one disk file per data set. Typically a software program is used to assist the entry of information so that the metadata conform to the standard.

Specifically:

- Define exactly what data packaging is to be documented.
- Assemble information about the data set.
- Create a digital file containing the metadata, properly arranged.
- Check the syntactical structure of the file. Modify the arrangement of information and repeat until the syntactical structure is correct.
- Review the content of the metadata, verifying that the information describes the subject data completely and correctly.

A digression on conformance and interoperability

The various metadata standards are truly *content* standards. They may not dictate the layout of metadata in computer files. Since the standard is so complex, this has the practical effect that *almost any metadata can be said to conceptually conform to the standard*; the file containing metadata need only contain the appropriate information, and that information need not be easily interpretable or accessible by a person or even a computer.

This rather broad notion of conformance is not very useful. Unfortunately it is rather common. To be truly useful, the metadata must be clearly comparable with other metadata, not only in a visual sense, but also to software that indexes, searches, and retrieves the documents over the Internet. To accomplish this, there are several encoding standards that specify the content of a metadata entry for exchange between computers, For real value, metadata must be both *parseable*, meaning machine-readable, and *interoperable*, meaning they work with software used in services such as the FGDC Clearinghouse through OpenGIS Catalogueue Services. Fortunately, the pending ISO 19115 metadata standard includes informative guidance on how the metadata could be structured as XML for validation and exchange.

Parseable

To *parse* information is to analyse it by disassembling it and recognising its components. Metadata that are parseable clearly separate the information associated with each element from that of other elements. Moreover, the element values are not only separated from one another but are clearly related to the corresponding element names, and the element names are clearly related to each other as they are in the standard.

In practice this means that your metadata are usually arranged in a hierarchy, just as the elements are in the standard, and they must use standard names for the elements as a way to identify the information contained in the element values.

Interoperable

To operate with metadata service software, your metadata must be readable by that software. Generally this means that they must be parseable and must identify the elements in the manner expected by the software.

There is a general consensus that metadata should be exchanged in Extensible Markup Language (XML) conforming to a Document Type Declaration (DTD). In the World Wide Web Consortium, there is progress on developing successor to the DTD, known as XML-Schema. Support for XML in parsing and presentation solutions is widespread on the Web and is presumed in current draft standards of the ISO TC 211 and OpenGIS specifications.

What software is available to create and validate metadata?

No tool can check the *accuracy* of metadata. Moreover, no tool can determine whether the metadata properly include elements designated by the Standard to be conditional, or 'mandatory if applicable.' Consequently, some level of human review is required. But human review should be simpler in those cases where the metadata is known to have the correct syntactical structure.

Software cannot be said to conform to the Standard. Only metadata records in a given encoding form can be said to conform or not. A program that claimed to conform to the Standard would have to be incapable of producing output that did not conform. Such a tool would have to anticipate all possible data sets. Instead, tools should assist you in entering your metadata, and the output records must be checked for both conformance and accuracy in separate steps. At best one can describe or anticipate *compatibility* testing among software components.

Issues in Implementation

Vocabularies, Gazetteers and Thesauri

When searching for information, the inquirer may not find any references based on the words used to describe the information sought. This problem can be overcome by use of a thesaurus. In the context of metadata and other electronic documents, a thesaurus is a tool for the organisation and retrieval of information in electronic materials. It allows data to be indexed and retrieved in a consistent manner. It permits the display of hierarchies of concepts and ideas, leading the user, whether as indexer or information seeker, to define his or her search in terms that are most likely to lead to the retrieval of relevant information.

For example, it will allow improved information retrieval by providing successful searching on synonyms - if the user enters the term "farming" the thesaurus will find the term "agriculture". Hierarchies of meaning can be shown - the term "Great Britain" may retrieve data indexed with that term but could also expand the search to retrieve data on England, Wales and Scotland which have been indexed under those three terms. The term "meals on wheels", although in a hierarchy of terms related to food, could also be linked to concepts relating to personal social services and to the different categories of recipients and a user can elect to follow and retrieve these related terms. Consistent searching for metadata will be achieved if all those who prepare metadata use the same thesaurus.

Minimum collaboration with users during the definition and implementation phases: a user-friendly focus is needed

For a non-professional user, finding the information wanted is very difficult. Even if 'Help' or 'Tutorial' can be found in some metadata services, it is not very easy to understand what to do and where to type. Efforts must be made to explain what to ask for and to develop user-friendly and multi-lingual interfaces. If it takes too much time to understand how to react to metadata services, users will not stay long and will immediately complain! A dictionary, multilingual thesauri or catalogueues with keywords, should be provided to users to ensure that the same vocabulary is used. One of the most important things is to develop services that are not technology dependent and technology driven. Projects must be done in collaboration with users (who must first be identified).

User-expected content

Given the complex metadata models deployed, we can be reasonably certain that the metadata that is now presented from catalogueue services is almost always more than is expected by end users. It seems that the current tendency is to propose a complex database approach that seems to be very 'data producer oriented'. One can imagine that users are more interested in examples and benefits on how to use the proposed data sets than a detailed description of its structure and content. This can be accomplished through special presentations of metadata.

It is important to separate the content of spatial metadata with its means of presentation. Through applications such as the Extensible Markup Language (XML), documents with extensive detail can be rendered through different style sheets from one content source into many presentation forms suitable to different audiences. Further work on developing presentation methodologies is required to simplify the burden of understanding metadata by all.

Metadata for applications

There is a tendency to adapt the metadata structure and content to applications, for example, electronic commerce or data management within an organisation. Metadata that is created to satisfy a real need, rather than because it is seen as something that should be done in the general interest, is more likely to be well-written and maintained.

The OpenGIS Consortium is developing metadata structures and fields to describe software interfaces, exposed as "services" for external use. This services metadata will help intelligent software, through brokers known as service catalogueues, to discover available services that could ultimately be chained together to form new composite operations. Services also have necessary links to data classes and instances. The OGC Web Mapping Testbed is documenting this interaction as a contribution to metadata in ISO 19115.

A geographic information product identification mechanism

There is no current mechanism to provide identification numbers (Id) to the different GI products produced and offered to users. This missing element is a very important issue for those who are implementing in parallel a metadata service and an e-commerce solution.

To make the e-commerce of GI a reality a study on how a GI numbering system could be organised and implemented and by whom should be made. This system could be similar to the ones used for other products, such as books. It would be extremely helpful if the Global Spatial Data Infrastructure activity could develop initial guidance on the technical and political issues involved in establishing a data product identifier system that will work globally on digital and non-digital geospatial information.

Incentives for metadata development

The impressive list of incentives which includes financial resources, knowledge and expertise, standard and tools provided by the FGDC (Federal Geographic Data Committee - http://www.fgdc.gov) to stimulate the creation and maintenance of metadata content and services within the concept of the US Clearinghouse appeared to be a key success factor of the US metadata initiative. It is important that national and regional governments evaluate, recognize, and provide such incentives to metadata builders and managers. Some have started – France, Canada, Australia, the United States and other countries develop and provide free software and to metadata builders. It is anticipated that the widespread adoption of the ISO 19115 metadata standard will further encourage the development of an international base of free and commercial tools around a common standard.

Envisage legislation for public sector metadata content

In countries where legislation is the main engine for creating new or adapting existing public sector activities, new laws may be needed to encourage or require the collection and distribution of standards-based metadata by the GI public sector and by commercial enterprises that collect geospatial data for the public sector.

Recommendations

- ***The Cookbook authors recommend that you don't invent your own standard. Adopt or build a national profile of the ISO 19115 metadata standard once it is released as Draft International Standard.***

Standards are very expensive to create and build implementations for. National standards should be adopted with the intention of supporting the ISO 19115 metadata content standard when it becomes available. This will provide the greatest interoperability rewards in a global environment.

- ***The Cookbook authors recommend that you prioritise your data.***

Begin by documenting those data sets that have current or anticipated future use, data sets that form the framework upon which others are based, and data sets that represent your organisation's largest commitment in terms of effort or cost. Framework layers and special, unique layers of great interest should be adequately documented for use within your organisation and by those on the outside. Of course, all published data warrant documentation this way, but through setting priorities you will know what work you have ahead of you.

- ***The Cookbook authors suggest collecting metadata a little at a time.***

For detailed metadata such as FGDC and ISO, an enormous amount of possible information can be collected. Although all fields are never filled in, it provides an opportunity to store specific properties in their correct location within the standard structure. This facilitates their storage and discovery in catalogueues (See Chapter 4). If certain types of metadata are collected during the data collection process as part of the current workflow, then many 20-second notes can amount to a substantial story later on. This type of information cannot be easily collected after the fact.

- ***The Cookbook authors recommend the development of a coordinated spatial data product identifier system for use globally***

The GSDI Technical Working Group with policy assistance from the Steering Committee should develop initial guidance on the technical and political issues involved in establishing a data product identifier system that will work globally on digital and non-digital geospatial information.

- ***The Cookbook authors suggest that research into a common thematic classification system for geospatial data be conducted by the Technical Working Group of the GSDI.***

Whereas ISO TC 211 is developing general specifications and methodologies, and the OpenGIS Consortium is building software interfaces, no convened global organisation is known to be co-ordinating a common classification system for geospatial data. As a result, the use of competing thematic thesauri make distributed search difficult.

References and Links

Chenez, Christian and Gaël Kermarrec, "On-going Metadata Initiatives in Europe", 1999, 5th EC-GIS Workshop, Stresa, ITALY http://www.eurogi.org/forum/5thgeo.html

Metadata Home Page, US Federal Geographic Data Committee http://www.fgdc.gov/metadata/metadata.html

Metadata Home Page Australia and New Zealand Land Information Council http://www.anzlic.org.au/metadata/metagrp.html

Metadata Home Page National Geospatial Data Framework, UK http://www.ngdf.org.uk/Metadata

Metadata Home Page European Spatial Metadata Infrastructure http://www.esmi.org

Chapter Four: Geospatial Data Catalogue -- Making data discoverable

Editor: Doug Nebert, US FGDC

Introduction

An increasing volume of information is now considered critical to everyday decision making in modern society -- a large portion of this information is essentially related to "place" in the context of position on the Earth. As more online information includes some geographic context, the ability to describe, organise, and access it has become increasingly difficult. The ability to discover and access geographic data resources for use in visualisation, planning, and decision support is a requirement to support societies at the local, regional, national, and international levels. Common solutions have been developed and will be described in this Chapter by evaluating organisational approaches, comparing definitions of community, identifying common architectural solutions, and sharing a base of techniques that are implemented in available non-commercial and commercial standards-based software.

This Chapter presents the concepts, current practices, and designs for geospatial data discovery. It is intended as a guide to those interested in the management, development, and implementation of compatible discovery services in environments where the cross-domain publication of geographic information is desired. Organisational issues and roles are presented that are critical to the understanding and maintenance of the services within a larger spatial data infrastructure. The principles described herein can be interpreted and applied in a range of information management conditions from non-digital collections of map information, through small digital catalogues, to integrated repositories of data and metadata. Relevant standards and software are identified for evaluation and application.

Context and Rationale

Although the Internet is becoming the world's largest repository of knowledge, its navigation is hindered by the lack of a surrogate and comprehensive catalogue. As a result, one is delivered tens of thousands of candidate documents in response to a reasonable query from today's search engines. Fortunately, geographic information frequently has *signatures* of location in the form of coordinates or place names and even may have a reference date or time associated with the data. These metadata provide a key to a solution that can and does operate in an international context.

The library has long formed the primary metaphor for accumulation and management of knowledge about people, places, and things. Since the construction of the ancient library in Alexandria, Egypt to its modern day equivalents, libraries have employed classification systems, specialisation, and discipline to information in all forms. A central feature in this virtual library – and a critical part to its navigation and use – is the catalogue. In the context of geospatial information management, we use the descriptions of geospatial data, or metadata, as described in Chapter 2 as the common vocabulary to frame the structured fields of information that we seek to manage and to use in search and retrieval. These metadata elements are stored and served through a user-accessible catalogue of geospatial information.

Support of a discovery and access service for geospatial information is known variously within the geospatial community as "catalogue services" (OpenGIS Consortium), "Spatial Data Directory" (Australian Spatial data Infrastructure), and "Clearinghouse" (U.S. FGDC). Although they have different names, the goals of discovering geospatial data through the metadata properties they report are the same. For the purpose of consistency within this document, these services will be referred to as "catalogue services." Further integration of these services with web mapping, live access to spatial data, and additional services can lead to exciting user environments in which data can be discovered, evaluated, fused, and used in problem-solving. Whereas this chapter will focus on finding spatial data and services, combination of the practices described here with those in other chapters can expand the capabilities of your spatial data infrastructure.

Distributed Catalogue Concepts

The Catalogue Gateway and its user interface allows a user to query distributed collections of geospatial information through their metadata descriptions. This geospatial information may take the form of “data” or of services available to interact with geospatial data, described with complementary forms of metadata. Figure 4.1 shows the basic interactions of various individuals or organisations involved in the advertising and discovery of spatial data. The boxes are identifiable components of the distributed catalogue service; the lines that connect the boxes illustrate a specific set of interactions described by the words next to the line.

A user interested in locating geospatial information uses a search user interface, fills out a search form, specifying queries for data with certain properties. The search request is passed to the Catalogue Gateway and poses the query of one or more registered catalogue servers. Each catalogue server manages a collection of metadata entries. Within the metadata entries there are instructions on how to access the spatial data being described. There are a variety of user interfaces available in this type of Catalogue search in various national and regional SDIs around the world. Interoperable search across international Catalogues can be achieved through use of a common descriptive vocabulary (metadata), a common search and retrieval protocol, and a registration system for servers of metadata collections.

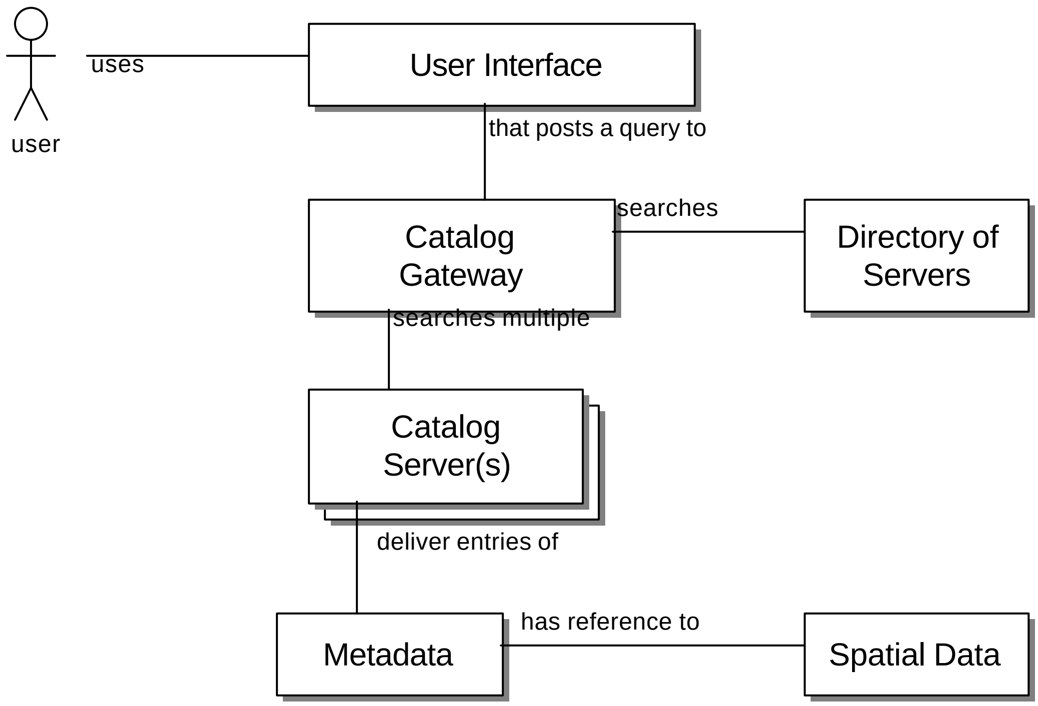


Figure 4.1 - Interaction diagram showing basic usage of distributed catalog services and related SDI elements from a user point of view.

The Distributed Catalogue environment is more than just a catalogue of locator records. The Distributed Catalogue includes reference and/or access to data, ordering mechanisms, map graphics for data browsing, and other detailed use information that are provided through the metadata entries. This metadata acts in three roles: 1) documenting the location of the information, 2) documenting the content and structures of the information, and 3) providing the end-user with detailed information on its appropriate use. A traditional catalogue, as found in the modern library, provides only locational information. In the era of digital data, the edges between the data or services and the catalogue can become blurred and permit the management of extended information called metadata that can be exploited by computer software and human eyes for many uses.

Organisational Approach

Who are the individuals or *actors* involved in the publication and discovery of geospatial information? By defining the roles and responsibilities that these actors play, one can understand the essential functions that human or computer-assisted services should be able to conduct in the interest of resource discovery for the GSDI.

Terminology:
Data Set – a specific packaging of geospatial information provided by a data producer or software, also known as a feature collection, image, or coverage.
Metadata - a formalised set of descriptive properties that is shared by a community to include guidance on expected structures, definitions, repeatability, and conditionality of elements.
Metadata Entry - a set of metadata that pertains specifically to a Data Set.
Catalogue - a single collection of Metadata Entries that is managed together.
Catalogue Service – a service that responds to requests for metadata in a Catalogue that comply with certain browse or search criteria.
Catalogue Entry - a single Metadata Entry made accessible through a Catalogue Service or stored in a Catalogue.
Service Entry – the metadata for an invokable service or operation, also known as operation or service metadata.

Roles

Figure 4.2 shows interactions between the Actors, the functions they perform, and the objects they interact with. The illustration uses Unified Modeling Language (UML) notation to picture processes from a functional point of view.

Originator of the Metadata Entry -- The responsibility of this Actor is to generate conformant metadata elements packaged so they accurately reflect the contents of the information being described. The role and credentials of the person responsible for the creation of this metadata may vary among organisations. In some situations the originator may be the scientist involved in building the data set being described. In others, the originator may be a contractor or second party who was directed to create the data or the metadata based on some project requirements, or it may be a generic description created by a production-oriented organisation without mention of the names of individuals involved in its creation. Given the rarity of metadata still, it is also a common practice for a third party to interpret or derive a metadata entry from available information where formal metadata has not yet been created.

Contributor to the Catalogue -- The responsibility of this Actor is to provide one or more conformant metadata entries to a Catalogue. Metadata entries may be delivered in proper format, derived from other formats, or developed from information stored in data and software systems. S/he interacts with the management functions of the Catalogue Service that permit a metadata entry to be entered, updated, deleted, or to assign levels of access or viewing privilege.

Catalogue Administrator -- The responsibility of the catalogue administrator is to manage the metadata for access by the Users. The maintainer or keeper of the metadata may be the same as the contributor, it may be a collecting organisation acting on the authority of an entire organisation (e.g. librarian or web site content manager), or it may be a different party who has acquired metadata in some form and is providing public access to it. The Custodian authorises access to the Catalogue Service for Management functions including entry, update, or deletion, manages authorisation details, and may perform some quality assurance evaluation on entries. The Custodian may also manage external (client) access to the Catalogue if it is not publicly accessible.

Catalogue User -- The responsibility of this user is to define criteria by which geographically-related information could be located and used through use of Browse categories or posing a fielded or full-text query. This user may or may not be GIS-literate, but with the Internet is likely to not be familiar with or possess GIS or image processing software. This User may have a weak understanding of geography. Another common method of catalogue access may be through a program to discover and work with Catalogue information. The interaction occurs at the software level and assumes a documented interface (e.g. application programming interface) for submitting requests to and receiving responses from a Catalogue.

Gateway Manager -- the responsibility of the manager is to develop, host, and maintain the distributed search capabilities within the user community. This may also include management of or contribution to a directory of servers (registry) that

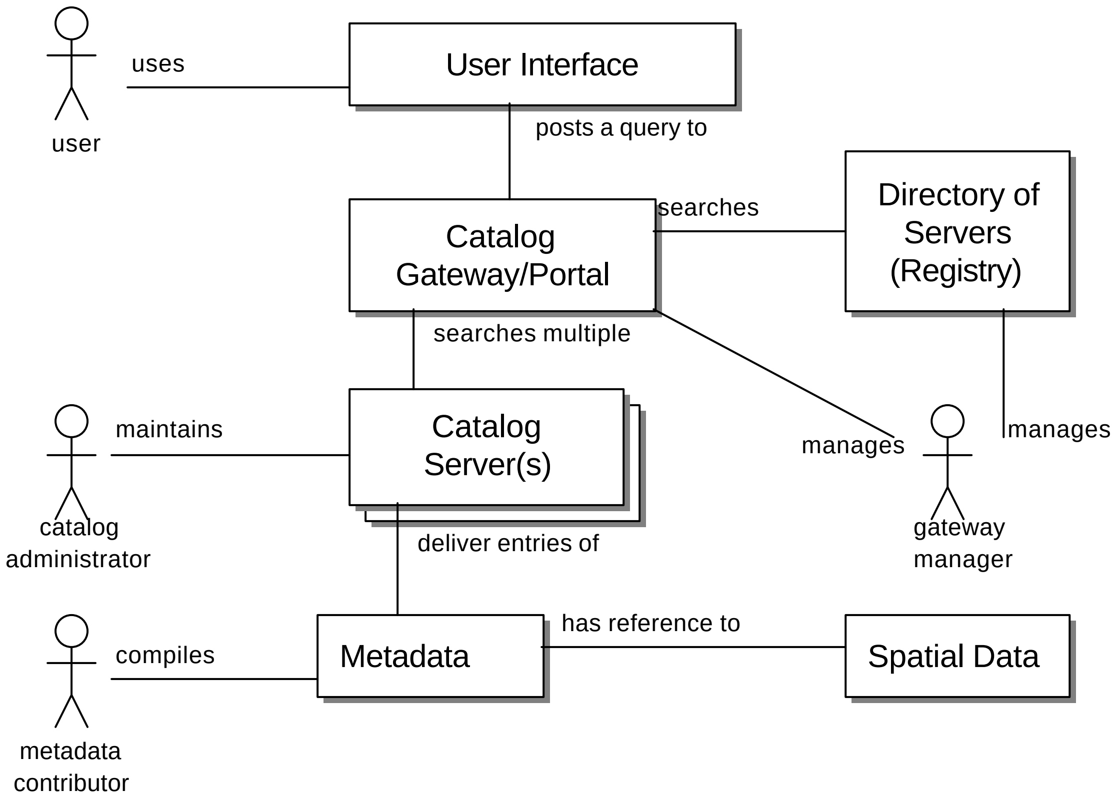


Figure 4.2 - Interaction diagram showing basic usage of catalog services and related SDI elements.

participate in the national or regional SDI.

Using the actors from Figure 4.2 as described in the text, the following sections will describe the organisational or operational management requirements for distributed catalogue services compatible with the GSDI based on the following areas of interest

- Catalogue Service development
- Catalogue gateway and access interfaces
- Registering participants

Each section will include a Use Case to focus on the roles and actions that should be considered in creating a discovery component of your SDI.

Catalogue Server/Service Development

The Distributed Catalogue services assume some degree of distributed ownership and participation. Similar activities on the Internet have taken a fully centralised approach to metadata management by placing all metadata in an index on one server, or in several replicated servers. In an increasingly dynamic data management environment, the synchronisation between detailed metadata and such an index becomes increasingly difficult. This problem is experienced on a daily basis when conducting searches on Web search engines and getting a "404: File not found" error when a document has been moved or changed. In addition we are seeing a migration toward treating metadata and data as interrelated and even being managed together within a single database. To duplicate this metadata in an external index can be costly and invites problems with synchronisation of the data, its metadata, and the externally indexed metadata. Organisations who already manage spatial data and are interested in publishing it are often the most capable candidates for publishing and maintaining the metadata. Metadata co-located with data on a server tend to be more current and detailed than metadata published to an external index (harvested and indexed off-site).

The construction of a catalogue service capability for geospatial information is built upon on the commitment to collect and manage some level of geospatial metadata within an organisation. The following Use Case scenario describes the publishing of a metadata entry.

A **contributor of metadata** receives the description of a new spatial data set developed by other professional staff. This metadata is generated in a transferable encoding format to allow exchange of the metadata without loss of context or information content.
This metadata entry is passed to a **catalogue administrator** for consideration and loading to the catalogue.
The **catalogue administrator** applies any acceptance criteria on the quality of the metadata as required by the organisation. If the metadata are acceptable it is inserted into the catalogue.

The **catalogue administrator** then updates the catalogue to reflect the new entry as available for public access.
This data set is now considered advertised because its metadata provide a searchable and browseable record of its background, its temporal and spatial extent, and many other searchable characteristics.

There are several models for where Catalogue services might be installed within or among organisations. Generally speaking, a catalogue server is usually installed at the level of organisation appropriate to the nature of the data or metadata, the organisational context or mandates, and the level at which a catalogue can be operationally supported.

Consortium Approach -- The consortium model is one where a single metadata catalogue is built and operated at one location and is shared by multiple organisations with a common discipline or geographic context. Metadata are exported from contributors and are forwarded to the common site where they may be evaluated, loaded, and made publicly accessible. This model may work well where there are personnel and computer access constraints and a shared service provides or extends outreach. The consortium approach also encourages collaboration between participants in building a collective data and metadata resource base across the organisations. The liabilities of this approach may include managing complexity and contributions from many sources and being sure that metadata provided stay synchronised with the data being described. Data might not be co-located with the catalogue service but may be referred to at contributor locations.

Corporate Approach -- The corporate model assumes that all metadata are forwarded within an organisation to a single service at which time corporate issues of quality, publication, style, and content may be evaluated. This model allows personnel and networking resources to be focused on developing and managing a single service and computer within an organisation. Some degree of policy must be established within the organisation for the collection and propagation of the metadata to the corporate host. This model is well-suited to organisations who may be restricted to providing a single public access computer for security reasons. The liabilities of this approach may include managing contributions from many sources within the organisation and being sure that metadata provided stay synchronised with the data being described. Data may be co-located with the catalogue service or may be referred to at contributor locations.

Workgroup Approach --The workgroup model assumes that a service would be established at each place within an organisation where data are collected, documented, managed, and served. This follows the trend on the Internet in which virtually anyone on a connected network can be considered a "publisher" of information. The workgroup model also assumes that the individuals and groups most closely associated with the collection and revision of the information are also involved in its catalogue and service. This can lead to a high degree of synchronisation between the data and their metadata -- in some cases, data and metadata warehouses could be completely integrated. The liabilities of this approach may include technical expertise in catalogues at the local level and coordination issues across a given organisation.

Because of the nature of the distributed catalogue and its ability to search many servers, all of the suggested models listed are equally viable. In fact, close reading of the model descriptions will show that they represent a continuum of organisational choices that vary in complexity, governance, and the degree of integration with the data being described.

Alternative Approaches

The operational design of a distributed catalogue as described above, depends in large part on the ability for clients to use the proposed services. Globally, access to computers and communications networks supporting Web applications is still available to a small minority of the population. While this is changing in almost all regions through providing community public access points, building and subsidizing network construction and interconnection, the distributed catalogue may not be well suited to conditions in many developed and developing countries where the Internet is not yet common or bandwidth is lacking. There are two solutions that have been prototyped and are suitable for public information access in such environments.

For organisations and clientele who have limited access to computers or networks, metadata can be reprocessed and printed and distributed as paper catalogues. Printing and distribution costs may be significant but a wide audience can be reached through public libraries and organisations interested in using spatial data in decision making. Synchronisation with current data content and holdings in such paper catalogues may also be an issue. Paper distribution of catalogues can always be considered a supplement to digital information service methods.

If Internet services are present and available to the public but network bandwidth within the region of interest is limited, individual catalogues may wish to support harvesting of metadata from remote sites in "mirror" catalogues. A good example of this would be for supporting regional data discovery across multiple servers in different locations whose connections are low-speed. If each catalogue posted its metadata in a Web-accessible directory, a crawler or harvester program could retrieve and index metadata from other sites into a regional or replicate index. This methodology is being demonstrated in the United States to provide a single synchronized point of access to metadata that are fetched from a small to moderate number of sites. Note that this still suggests that the combined collection itself is still behind a server with a common interface, but potentially fewer standing servers are required in this architecture. At the extreme end of this design one could envision a few large metadata repositories with common search interfaces. Primary concerns about the scalability of this approach include supporting extremely large searchable metadata indexes and the synchronization of the indexes with remotely held metadata and data. It is not likely that this approach would scale to support a single global collection of metadata using current technologies.

In environments where both data providers and clients have access to computers but not reliable networks, the creation of CD-ROM or DVD media with searchable metadata (and perhaps even data) is another outreach mechanism. Creation of digital media with metadata and data will be of greatest benefit where standard metadata and data approaches are followed, and a catalogue (software and data) could be placed on the media to minimise the cost of deployment where a catalogue already exists.

These alternatives should be viewed as approaches that supplement the catalogue services recommendations described in this Chapter until such time as the information is accessible to the majority of intended clients via the Internet. Use of the catalogue services will immediately enable international academic, commercial, and governmental use of such information for regional analysis issues.

Catalogue Gateway and Access Interface Development

Within a given geographic or discipline-based community, the need will exist to build relevant search capabilities that facilitate intuitive search across many servers. This problem can be divided into two related parts that must interrelate -- a user interface *(Search/Browse Interface, fig 4.2)* and a query distributor *(Catalogue/Gateway Portal, fig 4.2)*. When performed across the Internet, these functions may be logically deployed in different locations although they tend to be coupled together in server-

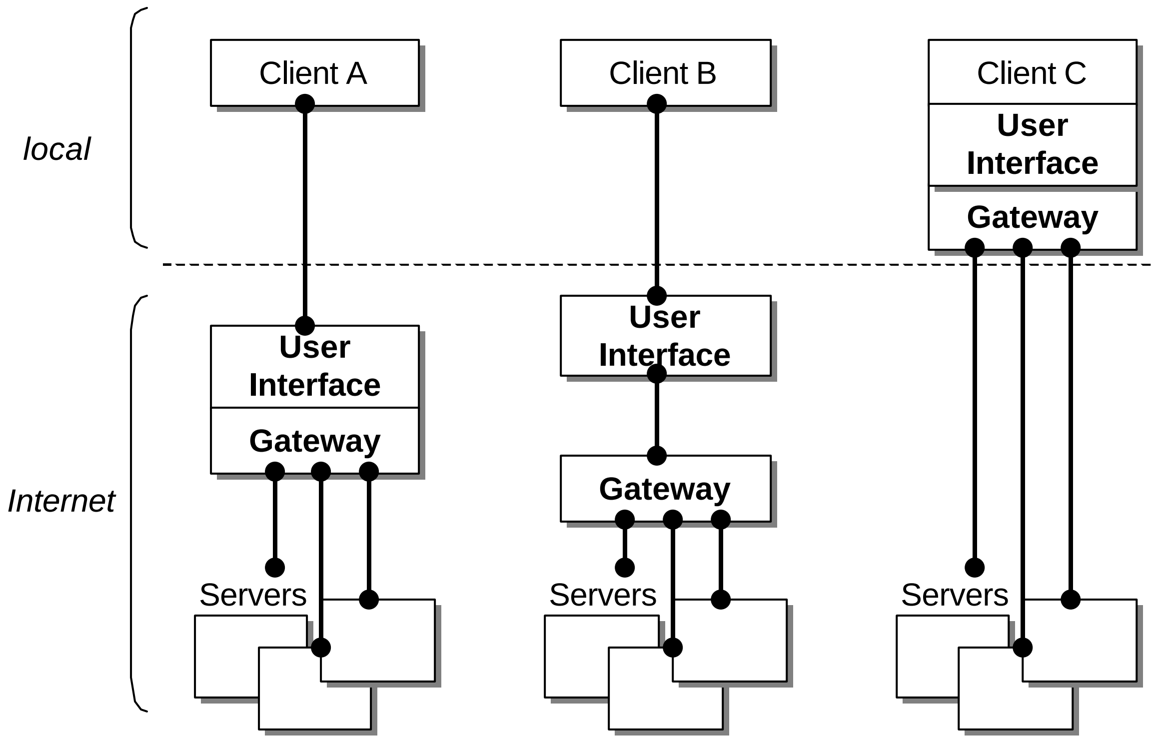


Figure 4.3 -- Configuration options for Gateway and User Interfaces to Distributed Catralog

based or client-based search solutions.

Figure 4.3 shows the possible configurations of a catalogue gateway and the user interface. Client A accesses a user interface that is downloaded (as forms or an applet) from a host on the Internet that is also managing multiple connections to servers. Client B is accessing a user interface from a location that is different from that of the Gateway supporting the construction of customised user interfaces for a community. Client C is a client-side "desktop" application that is fully self-contained and includes the user interface and distributed query capabilities for direct connection to remote servers. What is not known on this diagram is the dependence on or reference to a registry or Directory of Servers, as shown in Figure 4.2, which is further explained in the next section. All three styles of interaction are known to exist in various SDIs. Because they all depend upon distributed catalogue servers the three approaches are fully compatible.

Two styles of interaction are known to exist in Web search interfaces that are equally well applied to distributed catalogue access. The first style is **query** in which the user specifies search criteria for search using simple to advanced interfaces. The second style is a **browse** interface in which the user is presented with categories of information and selects paths or groupings, often in hierarchical form, to traverse.

The search approach to interaction with distributed catalogues can provide extra precision for advanced users in selecting spatial data of interest. It often is implemented in iteration to discover what effects individual parts of a query have on the pattern of results returned. The browse approach has great appeal to novice users who may wish to navigate by reference without knowing proper search words or fields *a priori*. The challenge of constructing and supporting browse mechanism across a global collection of servers is the work required in building and supporting a universal vocabulary for classification and its hierarchy or word space, known as an ontology. As this service lies at the intersection of many disciplines of interest, the construction of a single classification system is an extremely daunting and improbable task. Intelligent classification systems that are run externally on collections using neural networks, Bayesian probablitiies, and other estimates of "context" may be available in the coming years to help users navigate through heterogeneous geospatial information.

A Use Case scenario for a query user is as follows:

- A User uses client software to discover that a distributed catalogue search service exists.
- User opens the user interface and assembles the query elements required to narrow down a search of available information.
- The search request is passed to one or more servers based on user requirements through a gateway function. The search may be iterative, repeating or refining queries based on new interactions with the user.

- Results are returned from each server and are collated and presented to the User. Types of response styles may include: a list of "hits" in title and link format, a brief formatting of information, or a full presentation of metadata. Visualisation of multiple results may also be available through display of data set locations on a map, thematic groupings, or temporal extent.
- User selects the relevant metadata entry by name or reference and selects the presentation content (brief, full, other) and the format (HTML, XML, Text, other) for further review.
- User decides whether to acquire the data set through linkages in the metadata. By clicking on embedded Uniform Resource Locators (URLs) the user can directly access online ordering or downloadable resources, whereas distribution information lists alternate forms of access.

A User Case scenario for a browse user is as follows:

- A User uses client software to discover that a distributed catalogue search service exists. This may be done through a search of Web resources, a saved bookmark, reference from a referring page, or word-of-mouth referral.
- User opens the user interface and selects the parameters required to narrow down a search of available information based on topics/subjects, organisations, geographic bcation, or other criteria. Parameters are usually grouped into hierarchies for the user to navigate.
- Requests are made to each server through a distributed request mechanism.
- Results from each server are collated and presented to the User. Form of organisation of results is controlled by the user interface and gateway collaboration to present a uniform result space.
- User selects the relevant metadata entry by name or reference and selects the presentation content (brief, full, other) and the format (HTML, XML, Text, other) for further review.
- User decides whether to acquire the data set through linkages in the metadata. By clicking on embedded Uniform Resource Locators (URLs) the user can directly access online ordering or downloadable resources, whereas distribution information lists alternate forms of access.

Registering Catalogue Servers

The nature of distributed catalogues requires that the knowledge of the existence and properties of any given catalogue participating in a community be known to the community. In support of GSDI concepts, the need for a dynamic directory of services including catalogue servers is ever more important. The directory of servers concept allows an individual catalogue operator to construct and register service metadata with a central authority. This registry is then a searchable catalogue in its own right so that software may discover suitable catalogue targets based on their predominant geographic extent, descriptive words or classification, country of operation, or organisational affiliation, among other properties. Already national listings of compatible catalogue servers have been built, but the operation of a global network of catalogue servers within GSDI will require that a common directory of servers be built and managed to assure current content, distributed ownership, and authoritative reference to servers.

The features of the directory of servers may include:

- One descriptive entry per service collection (server metadata)
- Ability for a donor to contribute or update a record in the directory
- Ability to validate access to a server, as advertised
- User browse access of online server metadata
- Software search access of server metadata
- Management of active/inactive records, accessibility statistics

Several national distributed catalogue activities support management services for server-level metadata and contain references to servers predominantly in their country. The GSDI now sponsors a global directory of catalogue servers for all countries to utilise, with delegation of authority made to participating countries to manage and validate host information for their servers. This follows the Domain Name Service model of the Internet and if implemented in a similar way, would assure scalability and ownership within the global community. This registry can be accessed at http://registry.gsdi.org/registry .

Relevant Standards

The GSDI distributed catalogue has been designed with maximum reliance on existing technologies and standards. Because of this, existing software can be re-utilised or adapted to support geospatial information without requiring special investment in new technologies. Key standardisation efforts in access to catalogues are found in the ISO 23950 Search and Retrieve Protocol, the OpenGIS Consortium Catalogue Services Specification Version 1.0, and relevant standards or "recommendations" of the World Wide Web Consortium (W3C).

ISO 23950, also known as ANSI Z39.50, is a search and retrieval protocol developed initially in the library community for access to virtual catalogues. Key features of the ISO 23950 protocol include:

- Support of registered public "field" attributes for query across multiple servers where they may be mapped to private attributes
- Platform-independent implementation over TCP/IP using ASN.1 encoded protocol data units
- Ability to request both content (known as Element Sets or groups of ‘fields’ such as Brief or Full) and presentation format (Preferred Syntax, e.g. XML, HTML, text)
- GEO (Geospatial Metadata) Profile with registered implementation guidance for current FGDC and ANZLIC metadata and soon to include ISO 19115 metadata elements

The use of a generalised query protocol on ISO 23950 permits a migration from national forms of metadata to future forms being developed through international consensus under ISO Technical Committee 211 and their draft metadata standard 19115. Even though the metadata standard will change, the GEO Profile specifies the meaning of search fields in a way they can be mapped to multiple metadata schemas where compatible elements exist. Under the GEO Profile search of international metadata can be achieved today across collections in the United Kingdom (NGDF), the United States, Canada, Latin America, and Australia in a single search, even though different underlying metadata models exist.

The OpenGIS Consortium published a Catalogue Services Specification in 1999 that provides a general model for geospatial data discovery through a catalogue that includes management, discovery, and data access services. These general services are described for implementation in the OLEDB, CORBA, and WWW environments. The management functions include the ability to specify interfaces for creation, entry, update, and deletion of metadata entries to a catalogue. The discovery functions include the ability to search for and retrieve metadata entries from a catalogue with embedded references within the formal metadata to on-line data access, where available. The access functions support extended access to or ordering of spatial data based on references established in the metadata. Only the discovery functions are deemed mandatory in the Catalogue Services implementations; guidance is provided for implementation of optional management and access (really ordering) in interoperable ways.

At the August OGC meeting in Southampton, U.K., a common catalogue services approach was presented and demonstrated that built upon the essential search and retrieval model of ISO 23950. Implementation specifications were submitted for CORBA, OLEDB, and the Web. Distributed parallel search across these different protocols was demonstrated through an extension of commercially-available gateway software.

The Web Profile of the OGC Catalogue Services Specification includes two implementation paths: one permits the implementation of existing ISO 23950 servers (on TCP/IP) and a second specifies the use of XML encoding of queries and responses over HTTP. The XML Encoding Rules (XER) approach was demonstrated at Southampton through client and server software developed by the Joint Research Centre of the European. Because the server is implemented on HTTP, metadata providers need only install the server and index software as part of their web server as a component or module. Firewall issues of using a different TCP port are minimised because all queries could use the web server's communication port. OGC Web Mapping Testbed research has shown the popularity of an even simpler HTTP-based approach to catalogue services that still applies the basic tenets of ISO 23950. It is anticipated that the OGC Catalogue Services Specification Version 2.0 will include support for this approach as a new implementation Profile.

The International Standards Organisation (ISO) has a Technical Committee, TC 211, dedicated to the standardisation of abstract concepts relating to geospatial data, services, and the geomatics field in general. The draft metadata content standard (ISO 19115) provides a comprehensive vocabulary and structure of metadata that should be used to characterise geographic data. The development of national and discipline-oriented profiles of ISO 19115 will facilitate exchange of information using common semantics. ISO 19115 includes a recommendation on the encoding of metadata for exchange, using the encoding recommendation of TC 211, represented in XML format.

The World Wide Web Consortium (W3C) is a group of implementing organisations interested in developing common specifications, known as "recommendations' for wide support on the Web. One key set of recommendations and work items focus on the Extensible Markup Language (XML), a markup language specifically geared to encoding structured content of information. Companion topics include the XML-Schema activity, working on defining the schema and data types for XML documents and XML-Query -- at present only a design activity for a request syntax for XML-structured documents. The XML 1.0 Recommendation is in general use now, and is seeing wider application in the geographic software field as an increasingly richer means to encode and transfer structured information of all types. XML-Schema has recently been approved by the W3C and supports more rigorous validation of XML files.

Implementation Approach

The development of operational distributed catalogue services has been taking place in a number of countries including the United States, Canada, Mexico, Australia, and South Africa as primary examples. The software systems used to implement the ISO 23950 and Web based services has been developed largely through governmental support, resulting in both open source and commercial software solutions. The evolution of protocols and industry practices are difficult to predict, but this section provides a review of available solutions.

Let's review a technical use case scenario for access to a distributed catalogue:

- A User uses client software to discover that a distributed catalogue search service exists. This may be done through a search of Web resources, a saved bookmark, reference from a referring page, or word-of-mouth referral.
- User opens the user interface and assembles the parameters required to narrow down a search of available information.
- The search request is passed to one or more servers based on user requirements through a gateway service. The search may be iterative, repeating or refining queries based on new interactions with the user.
- Results are returned from each server and are collated and presented to the User. Types of response styles may include: a list of "hits" in title and link format, a brief formatting of information, or a full presentation of metadata. Visualisation of multiple results may also be available through display of data set locations on a map, thematic groupings, or temporal extent.

- User selects the relevant metadata entry by name or reference and selects the presentation content (brief, full, other) and the format (HTML, XML, Text, other) for further review.
- User decides whether to acquire the data set through linkages in the metadata. By clicking on embedded Uniform Resource Locators (URLs) the user can directly access online ordering or downloadable resources, whereas distribution information lists alternate forms of access.

The Distributed Catalogue is implemented using a multi-tier software architecture that includes a Client tier, a middleware or “Gateway” tier, and a server tier, as is illustrated in Figure 4.4.

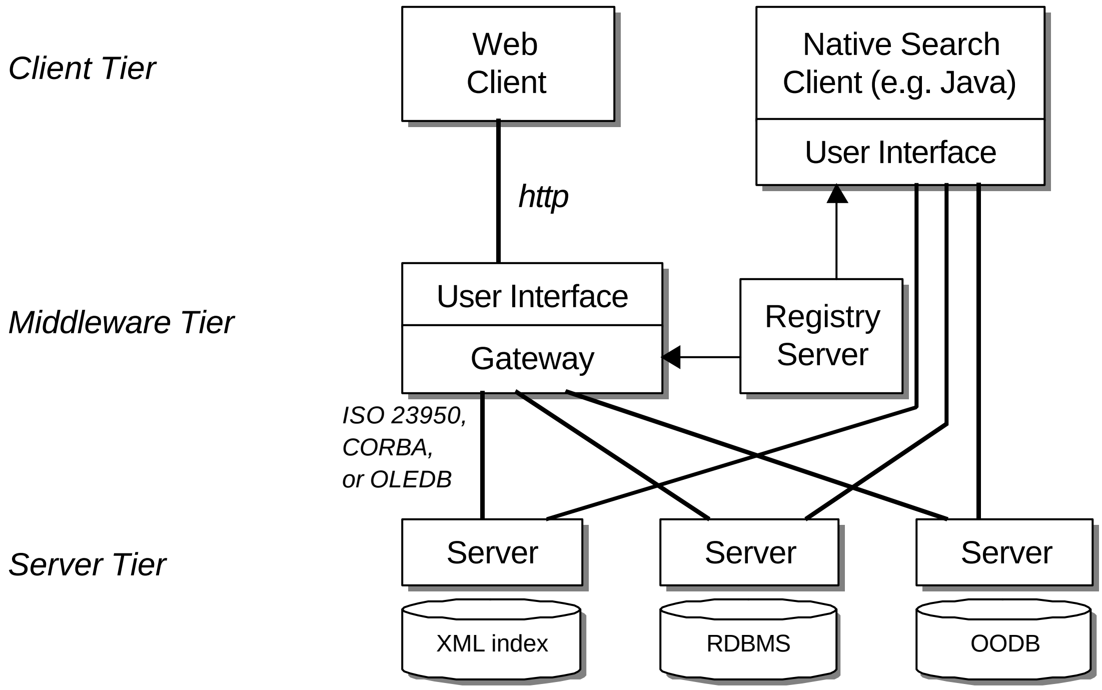


Figure 4.4 -- Implementation view of distributed catalog services

The client tier is realised by a traditional Web browser or a native search client application. The Web browser uses conventional HyperText Transport Protocol (HTTP) communications, whereas the native search client uses the ISO 23950 protocol directly against a set of servers. It is possible to also collapse this multi-tier architecture into two tiers where middle-tier functionality is present in the client.

The middle tier in the architecture includes a World Wide Web to catalogue services protocol gateway. A Gateway effectively converts an HTTP POST or GET request into multiple catalogue service clients that run either in series or in parallel. Gateway solutions provide parallel distributed search of multiple catalogue servers from a single client Web session. At present, Gateways have been installed in the U.S., Canada, Mexico, South Africa, Australia to provide regional points of access. The forms and interfaces installed at each are identical, and each hosts parallel search of all servers. In order to track a large number of Distributed Catalogue servers, a list of known, compatible servers called a Directory of Servers or Registry must also be managed. This service contains server or collection-level metadata that can itself be searched as a special catalogue. In this way, an intelligent one pass search of eligible servers can be performed instead of requiring the user to select servers from a list, or to have all queries passed to all servers.

At the bottom tier of the service architecture are the catalogue servers. These servers can be accessed using the GEO Profile of the ISO 23950 protocol, although CORBA implementations also exist. The GEO Profile of ISO 23950 is available to implementors in the geospatial community as an extended set of the traditional bibliographic fields that can be searched. GEO includes geospatial coordinates (latitude and longitude) and temporal fields in addition to free-text (e.g. search for the word anywhere in the metadata entry). ISO 23950 servers may be implemented on top of XML document databases, object-relational, or relational database systems in which structured metadata are stored for search and presentation.

The ISO 23950 protocol was selected for use in the Distributed Catalogue for several reasons. First, the library catalogue service community existed with relevant software and specifications that could be enhanced for geospatial search. By adopting compatible terms, library catalogues can be searched with GEO catalogues. Second, the ISO 23950 protocol specifies only client and search behavior and does not specify the native data structures or query language used to manage the metadata behind the server. Abstraction of query allows for a public query on “well known” fields that can be translated at each server into local equivalents. This lets one keep current database structures and names but supports alternative access through this geospatial public "view," expressed in XML or HTML reporting forms. This common search functionality across hundreds of servers is a prerequisite to distributed search. It allows for local database management autonomy yet supports federated search. Third, the protocol is independent of computer platform. ISO 23950 search clients and servers exist for many types of UNIX and Windows platforms, and Java libraries are available for additional client and server programming.

This separation between local and public metadata search fields has allowed for the ISO 23950 search of many different types of metadata collections that support the GEO Profile, even though they may not support the same metadata model. For example, The Australia and New Zealand Land Information Council (ANZLIC) metadata contains different tag names than FGDC metadata in the US. Through standard translation tables in the server, search against ANZLIC's "Data Set Name" field is associated with "Title" (the query labels this as attribute number 4) in the registered public fields. As a result, Australian catalogue servers can be searched through the FGDC Clearinghouse Gateways but return metadata records of a different structure. The same approach could be applied to other community metadata services, such as those employed by the Directory Interchange Format (DIF) files used in the space and global change disciplines or other metadata standards with similar content. Ideally, metadata formats should be delivered in such a structure that they could be converted or translated for consistent presentation, even if they come from different communities. The Extensible Markup Language (XML) and translator software is starting to enable the transformation of different XML documents in different schemas.

Catalogue Server/Service Development

To encourage widespread participation in the Clearinghouse, catalogue service software has been developed under direction of the FGDC and other coordination organisations around the world. Reference implementations of software exist to provide a free or low-cost example of metadata management and Distributed Catalogue service that can be quickly implemented. The software can also be used as reference by commercial developers to test anticipated functionality and interoperability and to develop value-added products.

A catalogue service that participates in a distributed catalogue should fulfill the following requirements:

- Support of a standard protocol (ISO 23950 preferred) for search and retrieval on an Internet-accessible server. When conformance testing for OGC Catalogue Services profiles is available, servers should be certified as OpenGIS-compliant (no conformance test methodology exists as of February 2000).
- Linkage to an indexed metadata management system that supports multi-field queries on text, numeric, and extended (e.g. "bounding box") data types, supports AND and OR constructs, and can return entries in a structured form that are or can be converted into a requested report in HTML, XML, and text. This may be a relational database, an object-relational database, or an XML database, or even a request to a remote catalogue to perform cascading catalogue services.
- Ability to translate public fields/attribute structures into names and structures used in the metadata management system using a national or international vocabulary (ISO 19115, when available)
- Ability to add, update, or delete metadata entries in the metadata management system

Available Software Implementations

The Isite software suite is a reference implementation of the Catalogue server that includes an XML document database and an ISO 23950 server supporting the GEO Profile for use on Windows and UNIX platforms. The U.S. Federal Geographic Data Committee is one of several sponsors that continue to support the development of this open-source software code. Isite supports document types conforming to the ANZLIC (Australia/New Zealand), Directory Interchange Format (DIF), Federal

Geographic Data Committee's (FGDC) Content Standard for Digital Geospatial Metadata and is used in a number of countries that support these content standards. As soon as ISO 19115 is available as Draft International Standard, support of an ISO document type can begin for Essential (core metadata) and Full Profiles within the Isite package.

Several commercial catalogue services supporting the OpenGIS Consortium Catalogue Services Specification Version 1.0 Web Profile via ISO 23950 are available on the market today. Links to known commercial solutions are posted on the Federal Geographic Data Committee web site (http://www.fgdc.gov/clearinghouse). When Version 2.0 of the OGC Catalogue Services specification is released and conformance testing methodologies are available, validated OGC-compliant software will also be listed from the OpenGIS web site (http://www.opengis.org).

Catalogue Gateway and Access Interface Development

As depicted in Figures 4.3 and 4.4, there is often a need for an intermediary to provide application integration for an end user. Known as "application servers" or middleware, these hosts allow for the storage, construction, and download of user interfaces to end users and communicate with multiple catalogue servers simultaneously -- a feat not supported by many web browsers due to security settings.

Software systems, such as application servers, that integrate catalogue search and other GIS and mapping functions benefit from the community development of software development kits (SDKs) based on standards. SDKs can provide client and server libraries for catalogue search and other services based on standard interfaces. Through component architecture, these SDKs expedite development of advanced software by combining appropriate pieces of software together as needed, reducing the need for a programmer to learn the intricacies of a given service.

A UNIX-based reference implementation gateway from the World Wide Web to multiple ISO 23950 targets is available for non-commercial use from IndexData in Denmark, known as ZAP (http://www.indexdata.dk). A perl-based programming client library to ISO 23950 is also available from the Joint Research Centre in Italy (http://perlz.jrc.it/download). A Java-based distributed search module to multiple ISO 23950 targets from common web servers is also being commissioned as open source software by the US FGDC as is a client-side Java library.

Registering Catalogue Servers

The operation of a growing network of distributed catalogue servers requires the management of server-level information in a central location. This registry server, shown in Figure 4.4, essentially houses server or collection-level metadata for search and retrieval and use in distributed query. In this way a search may be first made of the registry of servers to identify candidate servers to target the query, and as a broker, the registry returns the list of likely targets based on criteria such as geographic and temporal extent and other search limits. A registry facility greatly improves the scalability of a national, regional or global network of catalogues.

In the context of the GSDI, a coordinated registry of catalogue (and other) services is needed. If all catalogues were registered into a common and distributed registry akin to the way the Domain Name System (DNS) works, resolution of appropriate hosts of geospatial information globally will be enabled.

The GSDI hosts a global, seachable registry of catalogue servers using Isite fed by XML generated from an Access database. All geospatial catalogues conforming to FGDC, ISO, or ANZLIC metadata profiles should be registered here. This will be replaced with a conformant OpenGIS Catalogue solution supporting ISO metadata in the coming year (http://registry.gsdi.org/registry). A coordinated registry between the U.S. and Canada is proposed through an interagency agreement between the FGDC/GSDI Secretariat and Geomatics Canada as a model for other countries to follow in managing and cooridnating their own national catalogue entries with the global system.

Recommendations

- ***The Cookbook authors recommend that organisations publish their metadata using OpenGIS Consortium Catalogue Services Specification.***

The use of this specification, and in particular the Web Profile (ISO 23950), has increasing support from information locator activities on the Web. Existing reference implementation software allows organisations to participate at a very low cost; commercial implementations allow organisations to grow their collections and applications.

- ***The Cookbook authors recommend that participants register their catalogue servers at the GSDI coordinated registry for geospatial catalogue services.***

The operation of a GSDI service registry is not within the scope of an individual national organisation or consortium such as OpenGIS. The GSDI is the rightful host for a service registry and a policy forum for adjudication of the policies associated with such a registry. By placing ones catalogue references in such a system they can be discovered in a trans-national context.

References and Linkages

Catalog Services Specification Version 1.0, 1999, Open GIS Consortium, (http://www.opengis.org/techno/specs.htm#implementation)

Z39.50 International Standard Agency Home Page, (http://lcweb.loc.gov/z3950/agency/)

Chapter Five: Geospatial Data Visualisation -- Web Mapping

Contributors: Steve Blake, Australia; Frank Lochter, Germany; Allan Doyle, USA

Introduction

This chapter documents simple web mapping concepts and tools that enable the visualization of geospatial information from various organisations and servers across the World Wide Web. The linkages with Chapter 4 – Geospatial Data Catalogues, are also explored. Discussed are the current best practices related to on-line mapping, and the progress of the OpenGIS Consortium's (OGC) Interoperability Program[1] (IP) to realize the dream of true inter-operability and disseminating a web mapping specification for the vendors to adopt and promulgate.

Consider these desires:

- Do you want to view your information on a map online? Perhaps either as a simple (one map at a time) view or to overlay views from other sources together to produce a customized map product on your computer screen?

- Do you want to post a map layer from your in-house GIS or image processing system onto the Web for others to see? Do you want to provide views of your metadata so that your clients can picture the data or product you are responsible for?

If the answer to these questions is yes, then you are probably interested in Web Mapping.

Context and Rationale

The rise of the Internet and specifically the World Wide Web (WWW) has created expectations for ready access to geospatial information on the Web through a common web browser. Mapping on the Web includes the presentation of general purpose maps to display locations and geographic backdrops, as well as more sophisticated interactive and customizable mapping tools. The intention of online or Web Mapping is to portray spatial information quickly and easily for most users, requiring only map reading skills. Web mapping services can be discovered through online directories that serve both *spatial data* (through metadata) and *services*

---

[1] The OGC Interoperability Program began as the OGC Web Mapping Testbed or WMT. Since then it has expanded to encompass a number of activities and is often referred to as IP 2000 or IP 2001, etc. depending on which year the activity falls under.

information (see for example the *OGC Catalogue Services* draft specification). In fact, web mapping services are often used to assist users in geospatial search systems, showing geographic context and extent of relevant data against base map reference data.

Web Mapping implemented as a set of proprietary systems works fine as long as everyone you deal with both internally within your organisation and externally utilizes this same proprietary software. Because of this obvious particular limitation the Open GIS Consortium developed a non-proprietary web mapping approach based on the concept of interoperability. The topic of this chapter is not complex on-line GIS, but simple web mapping concepts and tools, i.e. part of a portrayal service to show spatial information on-line when the information originates from several discrete data/ map servers (commonly from different organisations).

Open GIS Web Mapping Activities

The sudden rise of web mapping over the last several years (cf. *GIS Online : Information Retrieval, Mapping, and the Internet* by Brandon Plewe - OnWord Press; ISBN: 1566901375) is demonstrated in the *interoperability* vision held by the Open GIS Consortium's *Interoperability Program* initiatives. In the OGC, expert GIS and web mapping technology users work with GIS software vendors, earth imaging vendors, database software vendors, integrators, computer vendors and other technology providers to reach agreement on the technical details of open web mapping interfaces that allow these systems to work together over the Web.

Consensus among vendors in the OGC's Web Mapping Testbed has created ways for vendors to write software that enables users to immediately overlay and operate on views of digital thematic map data from different online sources offered though different vendor software. The Web Mapping Testbed has delivered, among other specifications, a set of common interfaces for communicating a few basic commands/ parameters that enable automatic overlays. This set of interfaces is known as the OpenGIS® Web Map Server Interfaces Implementation Specification[2] and was developed by over 20 participating organisations.

The Web Map Server (WMS) specifications offer a way to enable the visual overlay of complex and distributed geographic information (maps) simultaneously, over the Internet. Additionally, other OGC specifications will enable the sharing of geoprocessing services, such as coordinate transformation, over the WWW. Software developers and integrators who develop web mapping software or who seek to integrate these capabilities into general purpose information systems can add these open web mapping interfaces to their software.

"Web Mapping" refers, at a minimum, to the following actions:

---

[2] The specification can be found at http://www.opengis.org/techno/specs/00-028.pdf

- A Client makes requests to one or more Service Registries (based on the OpenGIS *Catalogue Services Specification*) to discover URLs of Web Map Servers containing desired information.
- Service Registries return URLs and also information about methods by which the discovered information at each URL can be accessed.
- The client locates one or more servers containing the desired information, and invokes them simultaneously.
- As directed by the Client, each Map Server accesses the information requested from it, and renders it suitable for displaying as one or more layers in a map composed of many layers.
    - Map Servers provide the display-ready information to the Client (or Clients), which then display it. Clients may display information from many sources in a single window.

The OpenGIS Web Mapping Specifications address basic Web computing, image access, display, and manipulation capabilities. That is, they specify the request and response protocols for open Web-based client / map server interactions. The first of these specifications, described below, are the product of OGC's successful Web Mapping Testbed. They complement the already-available OpenGIS Specifications such as Simple Features and Catalogue Services, as well as ISO metadata standards to provide the foundation on which pending OpenGIS Specifications will build an increasingly robust open environment for Web mapping. Further interoperability initiatives (IP 2000 and IP2001) are defining Web Feature Servers, Web Coverage Servers, and extensions to the Web Map Servers that will allow a higher degree of control over the symbolization.

Today, the WMS 1.0 defines three main interfaces that support Web Mapping: ***GetMap***, ***GetCapabilities*** and ***GetFeatureInfo*****;** these were demonstrated at the conclusion of Phase 1 (May – September 1999) of the Web Mapping Testbed and were released to the public in April 2000. ***GetMap*** specifies map request parameters that allow multiple servers to produce different map layers for a single client. ***GetCapabilitie****s* explains what a map server can do (so integrators know what to ask for). ***GetFeatureInfo*** specifies how to ask for more information about web map features.

These interfaces provide a high level of abstraction that hides the "heavy lifting" in the Web Mapping scenario. The heavy lifting includes finding remote data store servers, requesting data from them in specifically defined structures, attaching symbols intelligently, changing coordinate systems, and returning information ready to be displayed at the client – all in a matter of seconds.

Servers conforming to OpenGIS WMS 1.0 will geo-enable Web sites and mobile devices for many new applications of geospatial technology. Consider any of the application domains listed below. Wherever the purchasers of the technology have chosen not to limit their users to a solution based on single vendor client/server pairs, these uses of geospatial data will depend on interfaces that conform to the OpenGIS Web Map Interface Specification

- Business siting, market research, and other business geographic applications
- Cable, microwave, and cellular transmission installation planning
- Civil Engineering
- Education/training, distance learning, multi-disciplinary research collaboration
- Electronic libraries, electronic museums and galleries
- Emergency road services and 911 emergency response systems
- Environmental monitoring, global and local
- Facilities management
- Global disaster/emergency/crisis management
- Health care: telemedicine, better/faster care for rural trauma victims, patient monitoring, etc.
- Intelligent vehicle highway systems (IVHS)
- Maintenance of one's information context and connection (personal logical network) as one moves through space, bridging media and modality; mapping electronic locations of addresses to their physical locations; using concepts of reach space, co-location, and near-by.
- Military applications: surveillance, planning, training, command/control, logistics, targeting
- Municipal public works maintenance and administration
- Natural resource discovery, exploitation, and management
- Navigation
- Precision farming (GPS-guided controlled delivery of nutrients and chemicals based on Earth imagery or automated GPS-located soil or crop sampling)
- Product distribution/warehousing optimization
- Public safety - fire and police departments
- Recreation: hiking, boating, etc.
- Science: climate research, agronomy, biology, ecology, geology, and others
- Security monitoring and intrusion response
- Special wayfinding for elderly and disabled
- Telecommunications network planning -- mobile communications
- Transportation planning
- Urban and regional planning
- Water resource management

There is a productive recent trend within the OGC to use Interoperability Initiatives like the Web Mapping Testbed to rapidly produce OpenGIS Specifications, as opposed to creating all of them through a traditional committee process. IP2000, completed in late 2000, focused on map authoring and publication, integrating graphical data and data elements (legends, symbolization, etc.), clients that can exploit XML-encoded information, further work on catalogue and discovery services, and work on transporting XML encoded data over the Internet.

Organisational Approach

Web based mapping provides the functionality to help discover and visualize spatial information referenced from Catalogue Service Systems. A Catalogue Service System (described in Chapter 4) is implemented through Internet-based software that allows users to inventory, advertise, and access metadata and associated geospatial information within a global framework of servers. Figure 5.1 shows one scenario of a client accessing a Catalogue (actually the catalogue implements a Service Registry) to discover data and web mapping services and then requesting and displaying maps from different servers.

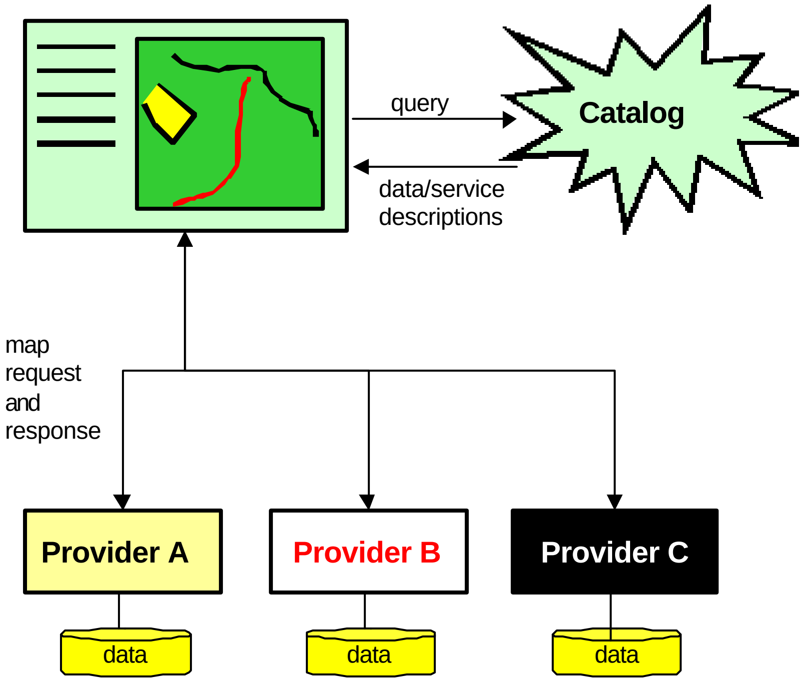
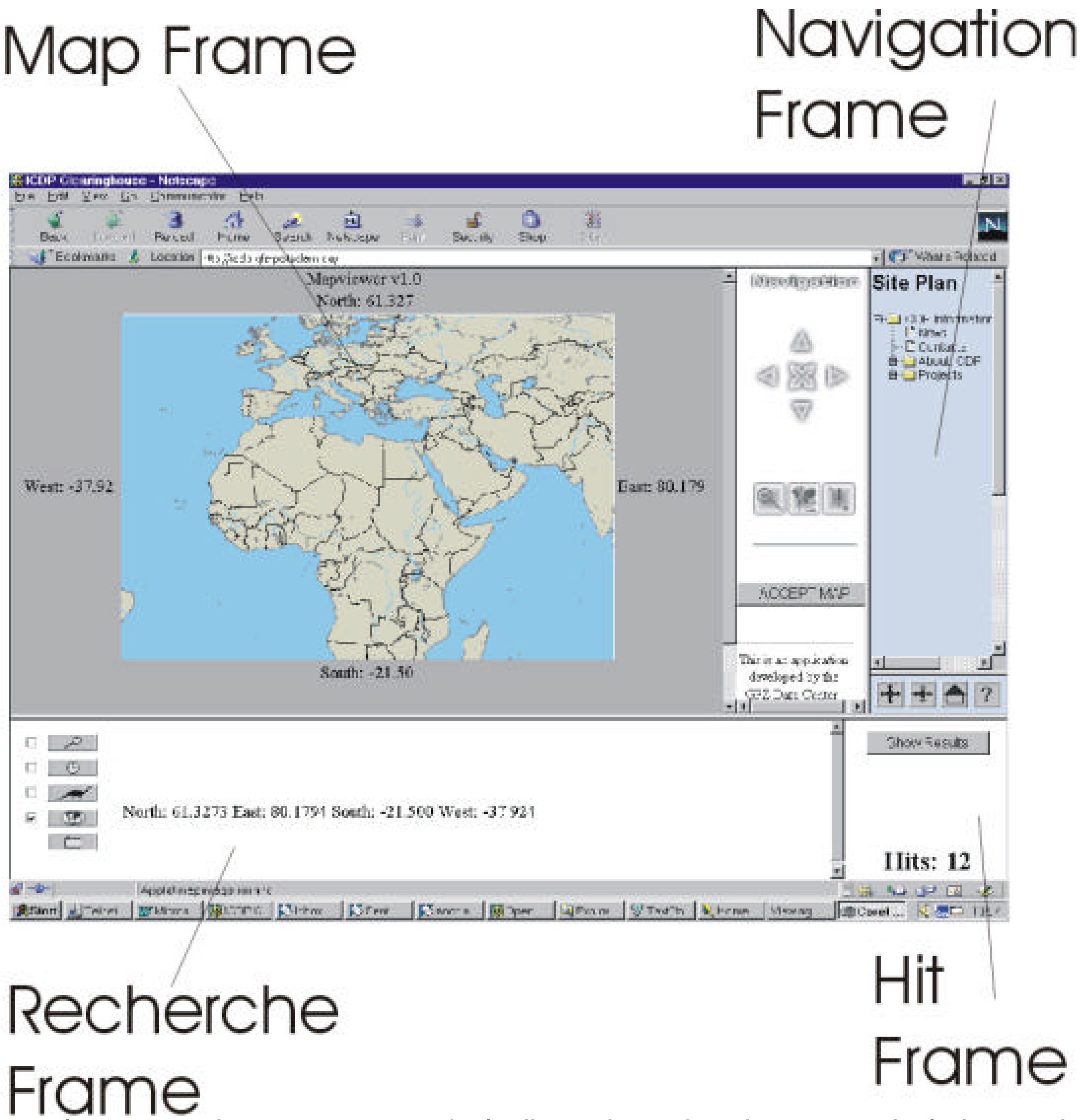


Figure 5.1 : Interaction of web map client with catalogue and map servers

A catalogue service that provides only references to raw geospatial data would be of use to only GIS experts and their software. By making map displays of geospatial information, casual users can interact with and see spatial data that was previously only available to GIS experts.

Figure 5.2 shows one example of a user interface for a Catalogue Service System. Many different GUIs can be built to provide special access for different categories of user. All the GUIs must use the same protocol agreements to interact with the map server software.

The Map Frame in figure 5.2 illustrates the value of specifying the bounding geometry (box or polygon) for the spatial part of the query for retrieval within the



Catalogue Service System. Typical dimensions for the query include spatial, temporal, paleotemporal and thematic values. The user also has the option to specify specific servers, or to search all registered servers for the geospatial data of interest.

The Map Frame can also be used for the presentation of the spatial component of the metadata in maps. The result presentation in a Catalogue Service System can be installed as a hidden search variable for further processing, or as List or Map in a web browser for visual presentation. The resulting presentation should be within the bounding geometry that was specified by the user for the Spatial Query. Often users like to interact with the objects on the maps. They like to have links on an object in a map connect to its metadata and then use a link in the metadata to connect to the real data. This can be accomplished via the GetFeatureInfo interface of the Web Map Server specification.

The success of Web Mapping depends on the use of consistent metadata standards (See Chapter 3). Historically, there have been a great variety of metadata standards developed and implemented across communities. Thanks to the contributions of many mapping organisations worldwide, an ISO standard 15046 for metadata is being finalized this year. Over time, organisations will see the value of migrating to a consistent ISO metadata format so that consistent global scale search and access of geospatial data can occur to support on line mapping.

Map Servers

For the concept of Web Mapping to be successful, a near global, truly inter-connected series of map servers must be established through the use of common protocols whether it be in an intranet, an extranet, or an internet scenario. Figure 5.3 provides a notional view of such a server network. Servers supporting on line web mapping will be registered to a Catalogue Service System as noted above.

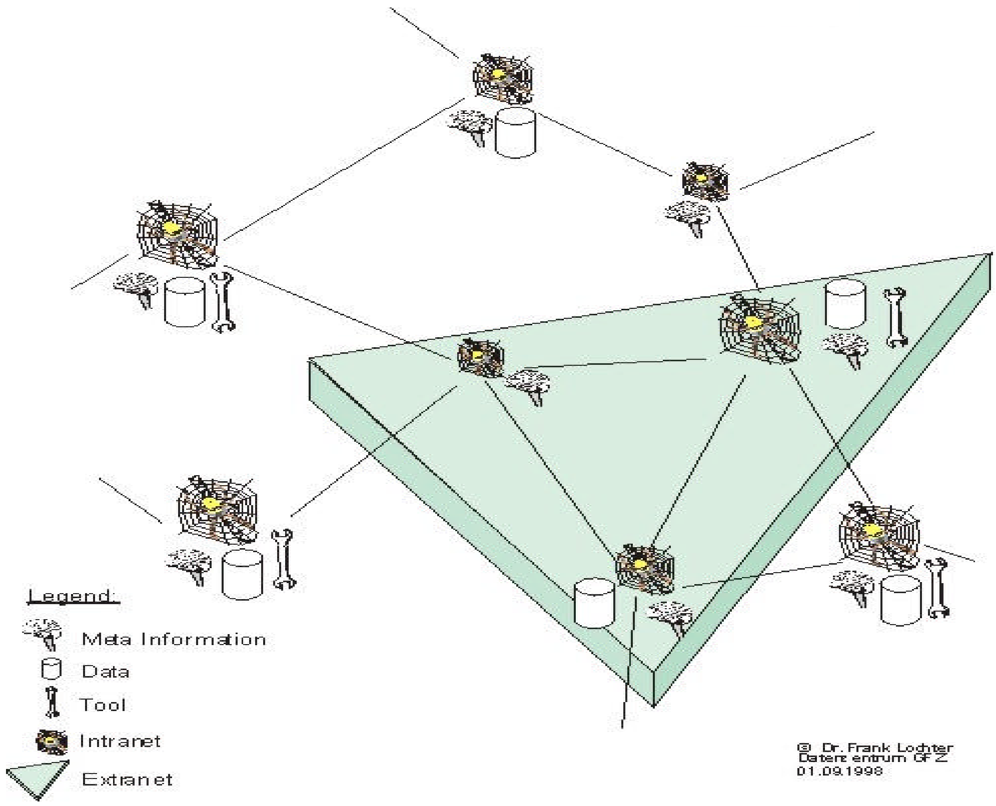


Figure 5.3: Notional View of Web Mapping Server Network

Implementation Approach

By way of introduction to implementations of Web Map Servers, the following is excerpted from the WMS 1.0 specification[3]:

A Map Server can do three things. It can:

- Produce a map (as a picture, as a series of graphical elements, or as a packaged set of geographic feature data),
- Answer basic queries about the content of the map, and
- Tell other programs what maps it can produce and which of those can be queried further.

To first order, a standard web browser can ask a Map Server to do these things just by submitting requests in the form of Uniform Resource Locators (URLs). The content of such URLs depends on which of the three tasks is requested. All URLs include a Web Mapping Technology specification version number and a request type parameter. In addition,

- To produce a map, the URL parameters indicate which portion of the Earth is to be mapped, the coordinate system to be used, the type(s) of information to be shown, the desired output format, and perhaps the output size, rendering style, or other parameters.
- To query the content of the map, the URL parameters indicate what map is being queried and which location on the map is of interest.
- To ask a Map Server about its holdings, the URL parameters includes the "capabilities" request type.

Each of these will be described in further detail later. We first provide some sample URLs and their resulting maps:

This requests a US National Oceanographic and Atmospheric Administration AVHRR image:

[3] The specification is under revision at the time of this publication, WMS 1.1 is expected to be published by September 2001.

```
http://a-map-co.com/mapserver.cgi?WMTVER=1.0.0&REQUEST=map&
SRS=EPSG%3A4326&BBOX=-97.105,24.913,78.794,36.358&
WIDTH=560&HEIGHT=350&LAYERS=AVHRR-09-27%3AMIT-mbay&STYLES=default&
FORMAT=PNG&BGCOLOR=0xFFFFFF&TRANSPARENT=TRUE&
EXCEPTIONS=INIMAGE&QUALITY=MEDIUM
```

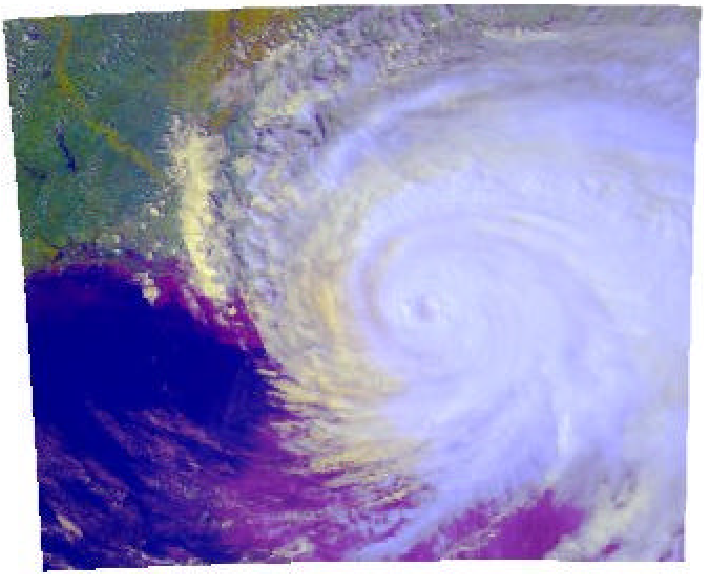


Figure 5.4 NOAA AVHRR Image of the Gulf of Mexico

This requests three layers, “built up areas”, political boundaries, and coastlines:

```
http://b-maps.com/map.cgi?WMTVER=1.0.0&REQUEST=map&
SRS=EPSG%3A4326&BBOX=-97.105,24.913,78.794,36.358&
WIDTH=560&HEIGHT=350&LAYERS=BUILTUPA_1M%3ACubeWerx,
COASTL_1M%3ACubeWerx,POLBNDL_1M%3ACubeWerx
&STYLES=0XFF8080,0X101040,BLACK&FORMAT=PNG&BGCOLOR=0xFFFFFF&
TRANSPARENT=FALSE&EXCEPTIONS=INIMAGE&QUALITY=MEDIUM
```

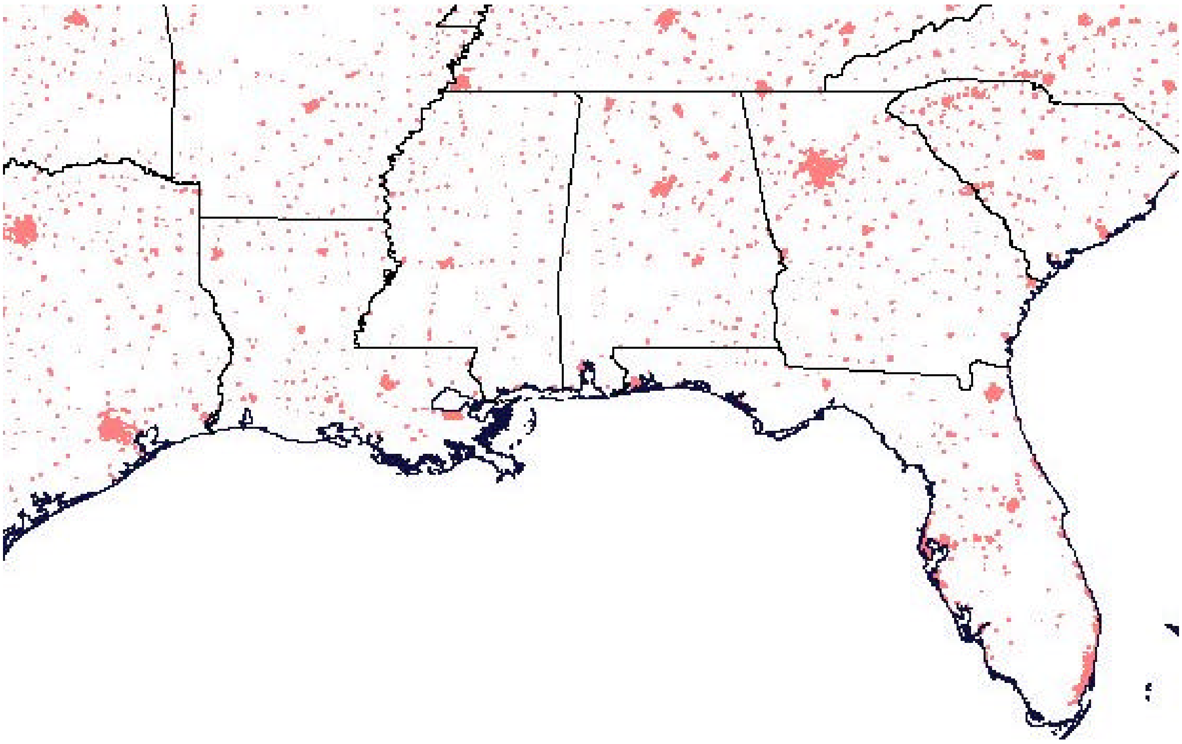


Figure 5.5 Political, Coastline, and Populated Areas, Southeastern United States

Notice that in both of these URLs the spatial information is identical:

SRS=EPSG%3A4326&BBOX=-97.105,24.913,78.794,36.358& WIDTH=560&HEIGHT=350

Because both maps were produced with the same bounding box, spatial reference system, and output size, the results can actually be overlaid by placing the latter map on top of the former. By enabling the use of image formats that provide for transparency information, maps that are meant to be overlaid over other maps can be produced by Map Servers. In this example, background areas of the second map are transparent (because the URL parameter "TRANSPARENT=TRUE" was supplied). Figure 5.6 shows the result of overlaying Figure 5.5 on top of Figure 5.4 to produce a map from the result of two separate Map requests. Finally, note that in this example the two maps were requested from different Map Servers. By standardizing the way in which maps are requested, clients of Map Servers can tailor which layers to request from which servers, thus building up maps that would not have been practical to assemble without the Web Mapping Interface Specification.

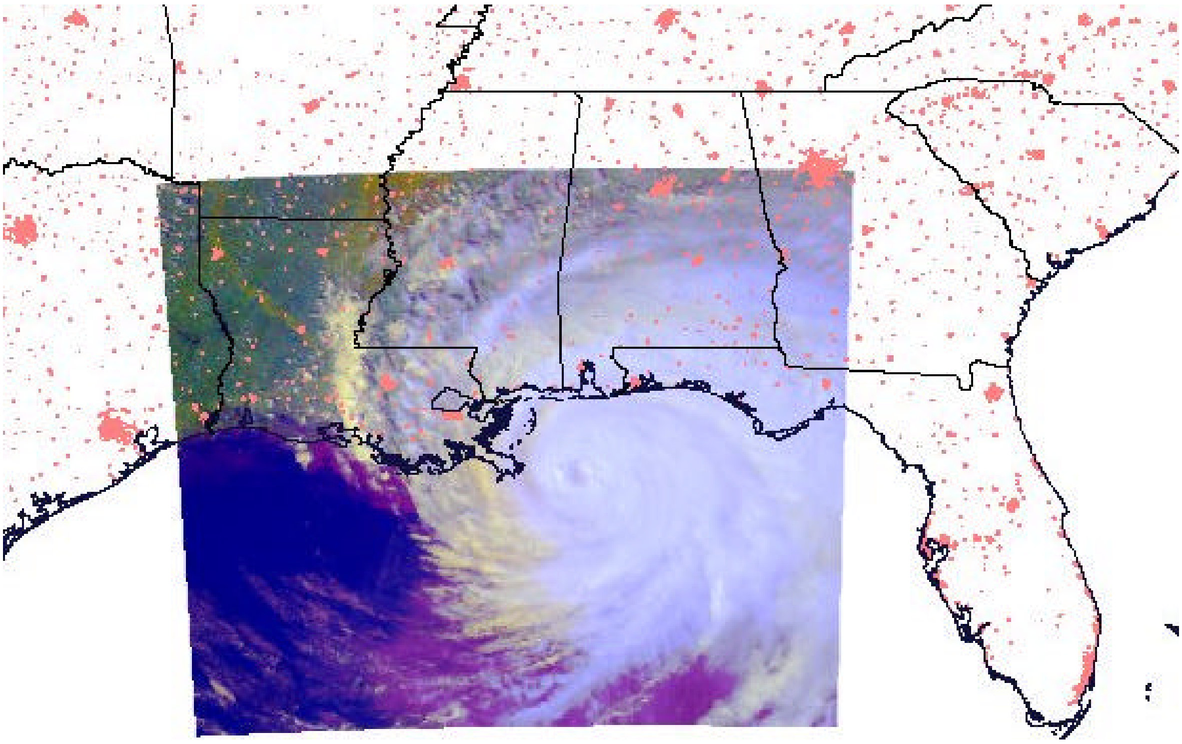


Figure 5.6 Combined AVHRR Image and Political/Cultural Map

If either of these maps were queryable, a client could request information about a feature on the map by adding to the map URL two additional parameters specifying a location (as an X, Y offset from the upper left corner).

Because each Map Server is likely to have different kinds of information for which it can produce maps, each Map Server must be able to provide a machine-parseable list of its capabilities. That enables the construction of searchable catalogues that can direct clients to particular Map Servers.

As a result of the OGC IP2000, additional interfaces that allow for the query of vector and raster information have been defined[4].

Since the completion of the first phase of the Web Mapping Testbed in August 1999, web mapping is fast becoming a reality. In the first phase of the Testbed participants developed the web mapping interface specifications discussed above. IP 2000 and

[4] Several interface definitions have been published as Open GIS Discussion Papers, these are meant to provide a view into the process but the content of the discussion papers is meant to be "unofficial" in the sense that they do not constitute Open GIS Specifications. Discussion papers can be found at http://www.opengis.org/techno/oip.htm#discussion

IP 2001[5], underway in 2000/2001, are evaluating and prototyping additional interface specifications for symbology, extended query and update, enhanced vector feature retrieval, and are further demonstrating interaction with the Catalogue Services specification. Additionally, future phases of the Interoperability Program may address e-commerce and security issues among other topics considered essential for robust implementation of web mapping.

Available Software

As a result of the Web Mapping Testbed, a number of GIS integrators and vendors have developed prototype versions of web mapping servers and compatible interfaces. The NASA-coordinated Digital Earth project includes software support for mapping NASA data using the specification (http://digitalearth.gsfc.nasa.gov/). OGC Web Mapping Service-compatible interfaces for ESRI Map Objects Internet Map Server version 1.0 and the University of Minnesota "mapserver" product (http://mapserver.gis.umn.edu) will be available to the public by May 2000 from participating organisations within the Web Mapping Testbed.

The US Army Corps of Engineers has developed a web site dedicated to Web Mapping at http://www.webmapping.org/

Other URLs of Web Map Servers built to the WMS 1.0 specification (as of August 14, 2000):

CubeWerx, Canada: http://www.cubewerx.com/wmt/

ESRI, USA
http://arconline.esri.com/arconline/downloads/ims_/WMS_Connector.html

Geodan, The Netherlands http://opengis.geodan.nl/opengis/wmt/

IONIC software, Belgium http://www.ionic.be/

SICAD Geomatics, Germany http://geo2.sicad.com:80/

Social Change Online, Australia
http://homer.socialchange.net.au/webmap/ogcwmt/demo/mapPage.jsp?appspec=applications%2Faus.xml&al_method=load&javaenabled=true

Demis, The Netherlands http://www2.demis.nl/mapserver/mapper.asp

European Commission Joint Research Center http://hgss.jrc.it/opengis/

---

[5] Information on IP2001 can be found at http://ip.opengis.org/ows/

Allan Doyle has written a WMS Cookbook to show how to build a very simple WMS, it can be found at
http://www.intl-interfaces.net/cookbook/WMS/

Recommendations

The state of Web Mapping is best illustrated by the progress made in the Open GIS Consortium Interoperability Program Activity. As the result of potentially competing vendors and software producers coming together and identifying a common set of functionality, a non-proprietary specification for rendering geo-referenced graphics has emerged. This allows one to establish a connection to multiple map servers and generate a stack of images that can be used in visual analysis and basic interrogation.

- ***The Cookbook authors recommend the use of the OpenGIS Web Mapping Services Specification, Version 1.0, or later***

Although further work needs to be done in the discovery, encoding, and exchange of 3-dimensional geospatial information in support of more advanced analysis and visualization, the basic Web Mapping Service functions provide an excellent starting point in the visual combination of distributed spatial data.

- ***The Cookbook authors invite all prospective organisations to participate in the development, prototyping, and establishment of next generation web mapping services in collaboration with the OpenGIS Consortium.***

References and Links

Digital Earth Reference Model Home Page, (http://digitalearth.gsfc.nasa.gov/RM/)

International Continental Scientific Drilling Program (ICDP). (http://icdp.gfz-potsdam.de/html/icdpinfonet.html)

Open GIS Consortium Web Mapping Testbed Public Page
(http://www.opengis.org/wmt/index.htm)

Open GIS Consortium Interoperability Program Page
(http://ip.opengis.org)

OpenGIS® Web Map Server Interfaces Implementation Specification Revision 1.0.0
(http://www.opengis.org/techno/specs/00-028.pdf)

Chapter Six: Geospatial Data Access and Delivery -- Open access to data

Editor: Brian McLeod, Canada

Context and Rationale

Access to geospatial data from the consumers point of view is a part of a process of that goes from discovery to evaluation, to access and finally to exploitation. Discovery (find, locate) involves the use of services such as metadata catalogues to find data of particular interest over a specific geographic region. Evaluation involves detailed reports, sample data and visualisation (e.g., in the recent form of web mapping through gifs or simple vector representations of the data) to help the consumer determine whether the data is of interest. 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 (use, employ) is what the consumer does with the data for their own purpose.

Typically in the past, the focus of geospatial data access was supplier side with a strong emphasis on technology and community based standards and specifications. With the growth of the Internet, in particular Web based technologies, access has become a demand driven operation. Consumers expect simple discover and access to cheap (or free) data in simple standard formats that can be used in desktop applications. Increasingly non-traditional suppliers are offering geospatial services, an example being Terraserver (http://www.terraserver.com/). The ability to leverage off other major developments such as the World Wide Web, and in some cases electronic commerce, has allowed broader participation in the Industry. The further democratisation of access to geospatial data thus enables value-added suppliers to create new data products and services.

The range of issues from an organisational point of view can be categorised two ways: 1) how broad is the client group; 2) how broad is the supplier group. In both cases issues tend to appear and grow as the groups become broader. In general issues revolve around copyright, licences (end user vs. reseller), cost, privacy, data formats and standards.

For example, if the client group is only internal staff then issues such as cost and copyright might not play a factor. As the scope of the client group grows to a limited number of known clients then there are straightforward mechanisms to control access. However, providing broad access to large group of potentially anonymous clients.

Similarly, as the size of the supplier group grows then issues appear. It is easier to establish a common policy for one or two organisations than it is for many. Typically each organisation has a business model (or non-business model!) that reflects its mandate and environment. The types of data and services it provides, the form and representation of the data, the quality and standards for the data all reflect this business model. Trying to bridge these issues between disparate organisations is an exponential problem.

The overlap between information managed by subject-specific communities in possibly parallel infrastructures can compound problems of data discovery and access. This can be viewed from either the consumer or supplier perspective. For example, as communities such biodiversity or geoscience specialists attempt to leverage a combined spatial data infrastructure to support their own goals they introduce new factors. These could be new standards or convention that they commonly require, it could be a new attribution requirement on the data not previously realised, or it could be the need to provide common access to data not otherwise visible from a spatial data infrastructure.

Several trends can be noted in the treatment and handling of geospatial data. Typically in the past the first concern of a data custodian has been what format the data is stored or managed in. Increasingly the trend is to move one level up and only worry about the interfaces to the data. This allows the data to be managed in the best manner possible, while providing open, standards based access. A consequence of this, however, is that the content of the data must be of a sufficient quality to support these interfaces. Often existing data is not accurate enough, up to date or lacking in attribution.

Another trend is in the organisation of the data itself. There is an evolution that starts back with traditional paper products. These migrated into discrete digital files that were typically stored offline, e.g., on a tape rack. As mass storage became more affordable these files found themselves living on online media (magnetic or optical) for easier access. This last step is an important one when you couple it with the developing of ubiquitous, wide area internetworking, i.e. the Internet. At this point a supplier was empowered to deliver data online.

More recently the trend has been to merge all the discrete data sets together into a single, seamless data warehouses, that have spawned the development of direct data access services. This has been enabled by developments in mass storage and spatial database technology. This step is also proving to be hard on the data, revealing inconsistencies in data accuracy and quality. Recent infrastructure developments allow the creation of virtual data warehouses that federate multiple instances of a data warehouses into a single logical entity.

Organisational Approach

As in any development it is important to understand who the stakeholders are and what roles each will play. For example in most national infrastructures government suppliers are key stakeholders. How they will play in the development and operation of the data access component of the infrastructure depends strongly on government policies regarding data distribution, cost recovery, etc.

Commercial entities will generally play a strong role as providers of tools and services but may also be suppliers of primary and value added data. It is important to understand the relationship between the commercial sector and the infrastructure as whole, e.g. will the commercial sector have a role in planning the infrastructure? What types of business arrangements will be supported in the infrastructure?

The final category of stakeholder is the consumer or end-user. Their use of the data access element infrastructure is dependent on a number of factors including: the functionality of the infrastructure tools, the amount and quality of the content accessible, operating policies, infrastructure business model (will consumers be charged for access?), etc..

In the early stages of the development it is important to specify and review the long term vision for the entire infrastructure to determine where the access components fits and how it ties into other infrastructure elements. At this stage it is helpful to develop some scenarios and use cases that can be presented to the stakeholders and refined as required.

The importance of developing a supportive policy/organisational environment should not be underestimated. Potential stakeholders will only become active participants if they see advantages for their organisations and if they do not feel threatened by the infrastructure. This policy/organisation environment will vary from country to country and will need to be worked out closely with the stakeholder community. The buy-in and commitment from senior management of all stakeholders is *critical* to the success of the infrastructure as a whole and to that of the access element in particular. The Canadian Geospatial Data Infrastructure **(**http://www.geoconnections.org/**)** is an example of an infrastructure implementation that has developed an organisation based on broad stakeholder participation.

Some of the issues that need to be considered in the development of the supportive policy/organisational environment are:

Distributed/autonomous suppliers

The management of the data should be done as close as possible to source. This ensures the accuracy and quality of the data.

Non threatening to mandates

Commercial and government stakeholders need to feel comfortable as active participants in the infrastructure. They should not feel threatened by infrastructure business models or policies.

Multiple levels of "buy-in"; low barrier to entry

The access component of the infrastructure must provide multiple levels of buyin from a low cost option with limited benefits, e.g. basic advertising of products and services, to higher cost options that offer increased benefits, e.g. distributed search connections to the supplier's inventory. This allows suppliers to choose a level of participation that best meets their business and operational objectives. This is especially important in the early operation of the access component as many suppliers will want to "try" it out and hence may not be prepared to expend much effort until they see how it works.

Sustainable long term business models

The access component of an infrastructure must provide an environment that supports a variety of supplier business models. The development of a sustainable business model for the operation of the access component is critical to the long term success of the entire infrastructure.

1. **Role of the private sector**
   The role the private sector as suppliers of data, services, and technology and as potential operators of the access infrastructure must be clearly defined.
2. **Marketing and promotion**
   The access component of an infrastructure must develop a marketing and promotion plan to build up the level of awareness and participation as quickly as possible. It is important to get a critical mass of suppliers so potential participants will see the benefits of joining the infrastructure. Potential benefits to suppliers include:

- Economies of data collection, closest to the source
- Reduced operational costs
- New clients (national and international)
- Data reuse (reuse vs recollection or conversion)
- Common tool and service reuse
- Advertising
- Benefits of “free” portrayal
- Enabling/supporting broad new applications, e.g.disaster management , value added

Implementation Approach

Definitions and Overview

Data Sets

Data sets are described by metadata and maintained within a data store. Foundation and Framework data sets represent fundamental data core that may be present within a spatial data infrastructure (See Chapter 2). Data sets are composed of collections of features (e.g. roads, rivers, political boundaries, etc.) and/or coverages (e.g. satellite/airborne imagery, digital elevation models, etc.).

Data Stores

Data stores are used to manage data sets. Data stores may be offline or online repositories. Traditional online data stores are file-based repositories, setup for the delivery of pre-defined data sets. Data stores also contain text and attribute data related to a data set. Data warehouses are datastores that provide seamless access and management of data sets.

Spatial Data Warehouse

A spatial data warehouse provides storage, management and direct access mechanisms. Typically, data warehouses ingest data from legacy file-based or data production systems.

Key characteristic of a spatial data warehouse include:

- the access and delivery of arbitrary features, layers, etc.
- seamless repository
- common data model
- application neutral, supporting a heterogeneous application environment
- support of large volumes of data
- multi-temporal support
- common repository for spatial and non-spatial data
- efficient access to large volumes of data

Commercial data warehouses include: Cubestore from Cubewerx (http://www.cubewerx.com/), the Oracle Spatial solution, (http://www.oracle.com/database/options/spatial/) and ESRI Spatial Data Engine (http://www.esri.com/) .

Data Access Service

Implementations of data access services include the following:

- Offline (e.g. packaging and physical delivery of data sets in either hardcopy or softcopy)
- Direct to datastore (e.g. softgoods delivery via ftp, specified via e-commerce order request)
- Brokered - provide specification of data access request to secondary (online or offline)access service
- Online data service (e.g. stateful request/response access protocol to data warehouse) supporting online operations such as:
- Drill down
- Aggregation
- Generalisation

In OpenGIS (http://www.opengis.org/) Project Document 98-060: "User Interaction with Geospatial Data" the Portrayal model is described. Figure 1 describes this model, which illustrates an simple features-based access and portrayal services pipeline.

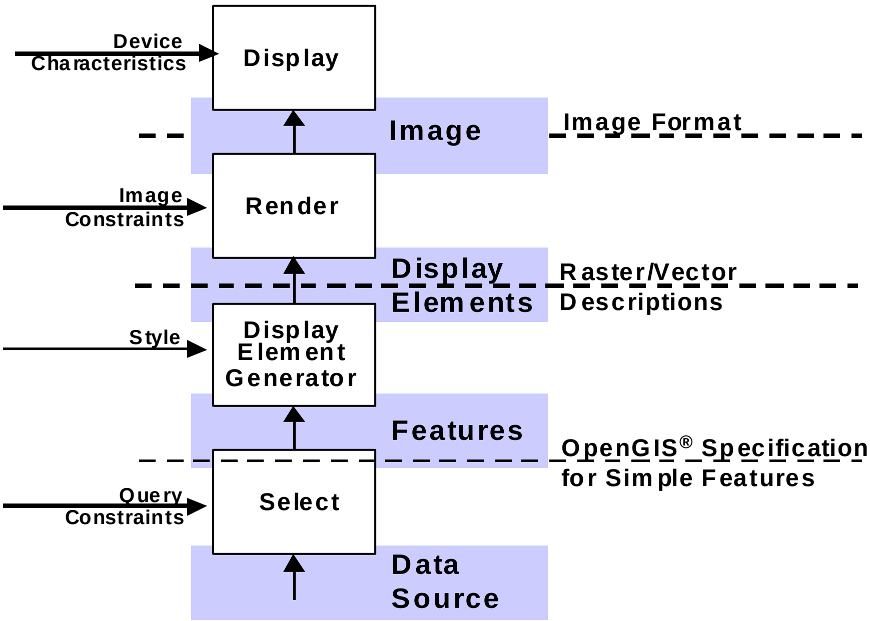


Figure 1- OGC portrayal model

Data Access Client

Online implementations of data access clients include:

- "thin" Internet/Web – client is provided by standard Internet/Web tools (no Java – e.g. Web browser, e-mail, ftp client, etc.)
- "medium" client provided by Web browser with Java, or ActiveX controls
- "thick" client provided by a Web browser plugin, or standalone application (network access via a distribution computing platform such as Corba, DCOM, Java RMI, etc.)
- Traditional GIS type client - access to previously downloaded data set, and direct network access to data warehouse
- "middleware" client – transparent access to consumer via a middleware infrastructure or applications service
- Geoprocessing service – direct access to data for use by a geoprocessing service (e.g. Web mapping [**ref: chapter 5**] interactive portrayal service)

Data Formats

Common spatial data formats include the following:

GIS proprietary (e.g. ESRI, MapInfo, Intergraph, etc.) A good overview of GIS formats can be found at http://www.gisdatadepot.com/helpdesk/formats.html

International and community Efforts have recently been made to minimise the number of geodata formats and to converge towards a reduced set. The Spatial Data Transfer System (SDTS), ISO TC/211 and the DIgital Geographic Exchange STandard (DIGEST) are examples of this trend. There are also exchange formats that allow the use of data outside of closed environments (e.g. Geography Markup Language - **http://www.opengis.org/**)

Typical native data formats for most GIS applications contain only enough information for the originating GIS application to be able to use it properly. The data formats usually carry the features and maybe some basic projection information. Data Exchange formats are usually more robust. They usually carry information that would allow the use of the data in a variety of systems. Exchange formats usually also carry some minimum metadata to describe the data set as well as data quality statements. Data exchange formats are typically used by producers of data. Due to lack of consensus on specific format standards, spatial data infrastructures must support for today's multitude of spatial data formats, and emerging data access services.

In the past, a multitude of GIS data formats were very problematic. Currently, most GIS and related access systems support format translation. Examples of commercial support for format translation include: the Feature Manipulation Engine from Safe Software (http://www.safe.com/) and Geogateway from PCI (http://www.pci.com/) An online data access service that combines data access with format translation is the Open Geospatial Datastore Interface (http://132.156.30.81/iii/).

Unfortunately format translation systems do little to support translation of semantics. The real problem for interoperable data access services, and formats is the lack of common semantics. Semantic translation and multi use feature coding catalogues (e.g. Digest) attempt to address the cross domain semantic support issue.

Web Implementation formats

Vector Files A vector file has many advantages that will prove useful for WWW spatial interfaces:

- A vector file can be delivered to the client where it can be zoomed and panned without the need to expensively conduct every operation on a WWW server.
- A vector file is composed of layers that might represent roads, rivers, boundaries.
- The layers can be switched on or off.
- A vector file allows a mechanism to limit the level of zoom so that spatial data is not pushed
- beyond its level of reliability.
- The size and efficiency of a simple vector file will help with network services and response times.

- Most GIS software can directly produce vector files.
- A vector file supports functions such as an interactive map.

There are a three candidate file formats for an inline vector file on the WWW: Simple Vector Format (http://www.w3.org/Graphics/SVG/), Web Computer Graphics Metafile (http://www.cgmopen.org/webcgmintro/paper.htm) and XML-based encoding formats (e.g. Geography Markup Language) which allows for Web-based transfer of feature information, for subsequent styling and rendering via Web client, or client plug-ins.

Raster Files Web/internet delivery of GIS raster formats such as ADRG, BIL and DEM (http://www.gisdatadepot.com/helpdesk/formats.html) is often problematic due to the large size of such files, combined with general lack of Internet bandwidth. Typically compressed raster files predominate Web-based portrayals for both Vector and Raster data. Common compressed Web formats include GIF, JPEG and PNG (http://www.cdrom.com/pub/png/).

Relationship to other spatial data infrastructure services

Figure 2 illustrates the relationship role of data access in an end-to-end resource discovery, evaluation and access paradigm. Successive iterations of resource discovery via a metadata catalogue, followed by resource evaluation (such as Web mapping) lead to data access either: direct as a data set, or indirect via a data access service.

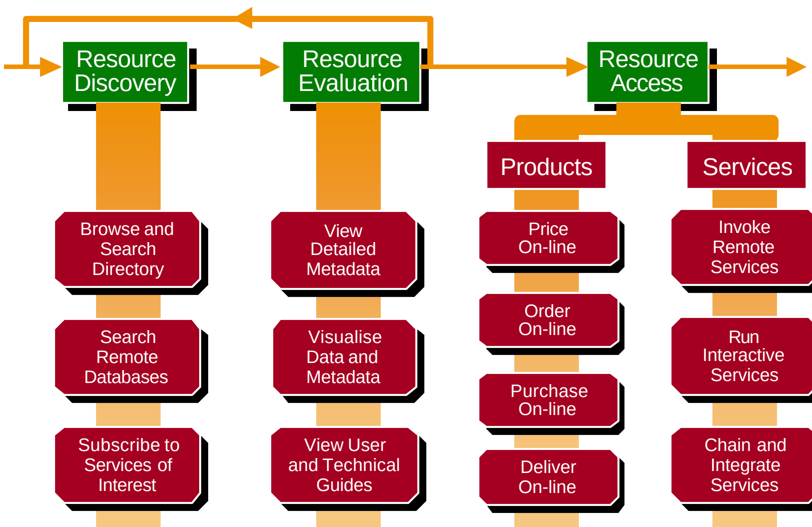


Figure 2 – Geospatial Resource Access Paradigm

Mature spatial data infrastructure will allow both application and human exploitation of the resource access paradigm. A key element of future spatial data infrastructures is the ability to broker requests for services, based on discovery and real-time access to online geoprocessing and related services. Future capability for chaining of distributed geoprocessing services is also expected.

A system context for data access is given in Figure 3. A data access service provides network access to a data set stored within a data store. Data sets are discovered (and later accessed) via metadata queries from a catalogue client to a data catalogue service [See Chapter 4].

Data sets can be visualised (and later accessed) via Web Mapping services [See Chapter 5], which are complementary to the data catalogue service.

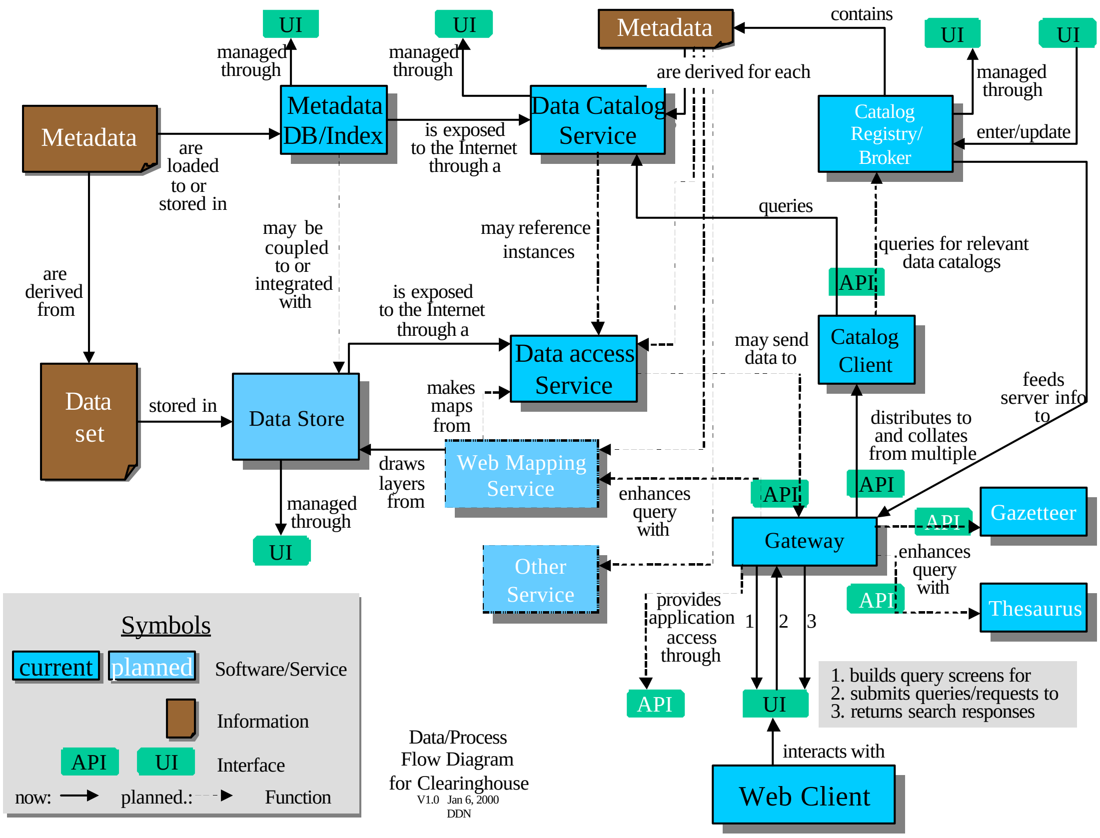


Figure 3 – System context for Geospatial data access services

Standards

In general, standards related to geospatial data access are still in their infancy. The standards of most relevance to access components of spatial data infrastructures include those from ISO/TC211, Open GIS Consortium (OGC) and Internet-related bodies including the World Wide Web consortium (W3C) and the Internet Engineering Task Force (IETF).

ISO/TC211

The primary mandate of ISO/TC211 (http://www.statkart.no/isotc211) is international standardisation in the field of digital geographic information.

"This work aims to establish a structured set of standards for information concerning objects or phenomena that are directly or indirectly associated with a location relative to the Earth.

These standards may specify, for geographic information, methods, tools and services for data management (including definition and description), acquiring, processing, analyzing, accessing, presenting and transferring such data in digital/electronic form between different users, systems and locations.

The work shall link to appropriate standards for information technology and data where possible, and provide a framework for the development of sector-specific applications using geographic data."

Emerging work on services is currently underway in both ISO/TC211 and the OGC. The definition of services interfaces will allow a wide range of applications access and use of geospatial resources. The OGC Simple Features Access model for SQL has been submitted to ISO for standardisation.

ISO SQL/MM

The purpose of the Draft Spatial Database Standard SQL/MultiMedia (SQL/MM) Part Three Spatial is to define multimedia and application specific objects and their associated methods (object packages) using the object-oriented features in SQL3 (ISO/IEC Project 1.21.3.4).

SQL/MM is structured as a multi-part standard. It consists of the following parts:

- Part 1: Framework
- Part 2: Full-Text
- Part 3: Spatial
- Part 4: General Purpose Facilities
- Part 5: Still Image

SQL/MM Part 3: Spatial is aimed at providing database capabilities to facilitate increased interoperability and more robust management of spatial data.

Open GIS Consortium (OGC)

Phase 1 of the recent OGC (http://www.opengis.org/) sponsored Web Mapping Test (WMT) bed initiative [ref: Chapter 5] has been successful in "Webmapping" portrayal of spatial data. An XML-based encoding scheme (Geography Markup Language or GML) for OGC Simple features was also an important output of WMT phase 1. Further evolution of the GML specification and direct data access is expected in subsequent OGC testbed initiatives, including a WMT phase 2.

Other activities of the OGC include the following:

The Open GIS Consortium has achieved consensus on several families of interfaces, and some of these have now been implemented in Off-The-Shelf software. All OGC consensus interface specifications carry a pledge of commercial or community implementation by their submitting teams.

Three Open GIS Simple Feature Access (SFA) interface specifications have been released: one each for SQL, COM-based, and CORBA distributed computing platforms. Companies belonging to the teams submitting one or more of these include Bentley Systems, ESRI, Oracle, Sun Microsystems, UCLA, Camber, Intergraph, Laser-Scan, MapInfo, Smallworldwide, IBM, and Informix.

The interfaces provide several layers of access to and control over GIS features. At the primitive level, the interfaces provide for the establishment of linear and angular units, spheroids, datums, prime meridians, and map projections that give semantics to coordinates. At the intermediate level, they enable the construction and manipulation of geometric elements such as points, lines, curves, strings, rings, polygons, and surfaces, as well as the topological and geometric and other relationships between them. Included are support for common geometric and topological constructs, such as convex hull, symmetric difference, closure, intersection, buffer, length, distance, and dozens of others.

At the GIS feature level, the interfaces provide for the creation and management of feature collections, and the ability to access features from such collections using geometric, topological, or attributional modifiers. Features and feature collections may be invoked in Well-Known-Binary (WKB) or Well-Known-Text (WKT) codes. Work is underway to specify Simple Features Access encoding using the Extended Markup Language (XML) as a well-known packaging of geometric and attribute information.

Open Geospatial Datastore Interface (OGDI)

OGDI offers a data access approach that leverages and accelerates standardisation efforts. OGDI is an application programming interface (API) that resides between an application and various geodata products, to provide standardised geospatial access method. The publicly available OGDI specification, and reference implementations are maintained by the Internet Interoperability Institute (http://132.156.30.81/iii/).

OGDI uses a client/server architecture to facilitate the dissemination of geodata products over the Internet/Intranet and a driver-oriented approach to facilitate access to a variety of geodata products and formats.

OGDI features include the following:

- **the distribution of geodata products via Internet/Intranet**. This reduces the space needed to store geographic data and insures access to "closest to the source", up-to-date data.
- **access to data in native format**. There is no need to keep multiple versions of geographic data in order to accommodate different GIS software packages.
- the adjustment of coordinate systems and cartographic projections; done on-the-fly so that original data is unaltered.
- the retrieval of geometric and attribute data.
- access to a large number of geodata products and formats.

Web and Internet related

The Internet Engineering task force (http://www.ietf.org/) develops and maintains specification for many Internet related application, transport, routing and security standards (Request for Comments – RFCs) many of which are related to data access (e.g. http, ftp, smtp)

The World Wide Web consortium, or W3C (http://www.w3.org/) is responsible for the development of common protocols and specifications to further the evolution of the World Wide Web. Activities of the W3C that related to spatial data access include work on Web graphic file formats, XML and metadata.

Related Services

Many services are related to data access. A brief listing follows:

Discovery and catalogue services [**ref Chapter 4**]
Webmapping [**ref Chapter 5**]
Electronic commerce related (e.g. http://www.commerce.net/)
Authentication
Payment
Confidentiality (e.g. Secure Socket Layer)
Public Key Infrastructure
Delivery and Packaging
Compression
Subsetting and subselection
Container-based delivery systems (e.g. http://www.paradata.com/)
Data subscription services
Data and file transport

HTTP
FTP
SMTP/MIME
Geoprocessing services (e.g. as defined by OGC)
Distributed Computing Platforms
CORBA (http://www.omg.org/)
COM (http://www.microsoft.com/)
Web/Java/XML

Best Practice Application

GeoGratis (http://geogratis.cgdi.gc.ca/)

One common problem with online access to data through a single infrastructure is the variety of policies and practice in place by the different data custodians. In order to support these different access policies one approach is to develop services to support different basic paradigms. These cases include:

- Custodians who restrict access to particular users would benefit from common user authentication/authorisation services;
- Custodians who charge for data or services would benefit from electronic commerce services;
- Custodians who distribute data free of charge would benefit from an inexpensive mechanism (both time and money) to distribute data.

One example of services to support the third paradigm is GeoGratis that provides common services to support the distribution of freely available geospatial data. GeoGratis provides a single ftp/web access point where consumers can discover and download freely available data sets. As a common online service GeoGratis can be viewed from different perspectives:

- The types of data it makes available;
- The services it provides;
- The distribution model it offers.

GeoGratis makes many types of geospatial data available to the consumer. These data may be national or local in scope, raster or vector, or current or legacy data.

Small-scale national data sets are commonly made publicly available. In the case of GeoGratis, base map data from the National Atlas of Canada is available for download. Additionally many national scale framework data sets are available through GeoGratis. At the other end of the spectrum are data from local test studies/sites that are nominally available free of charge. By offering basic download capabilities GeoGratis supports a wide variety of data types, including raster, vector and tabular. The only restriction is on any value-added service above the basic download capability. A final characteristic of the data available through GeoGratis is the availability of many legacy data sets such as the Canada Land Inventory. These data are typically data sets that suffered through some measure of cost cutting or program termination and as a result are no longer supported. GeoGratis provides a facility to make these data available albeit without background support.

In addition to freely available data GeoGratis provides value-added services. As a basic service GeoGratis provides the download of freely available data. Other basic services that GeoGratis provides is the discovery of available data through a search interface, the evaluation of data sets through detailed metadata and visualisation. Additionally, extra services are provided in support of data download – these include data subsetting, reprojection and reformating for all types of data available through GeoGratis. More advance services include the provision of data warehousing capabilities that support seamless access to large area data sets available through GeoGratis

Finally, GeoGratis offers a cost avoidance data distribution model. Since GeoGratis is provided as one of many common services supporting data access, this distribution model does not preclude other models, i.e., private access or fee based access. Similarly, GeoGratis does assert that all data should be freely available, but provides an effective service for data that is freely available.

One example of this is the National Atlas of Canada digital data. Originally these data were sold for a nominal fee. However it did not prove cost effective to continue this strategy due to the costs of selling and supporting the data compared to the limited return. Therefore a strategy of cost avoidance was adopted where the data was placed on GeoGratis for free download and support was removed. Access by any other means (such as distribution of the data on CD) was left to the value added private sector community. The result was a dramatic increase in the access and use of these data.

From an implementation and standards perspective, Geogratis provides an excellent "data rich" environment in which to implement emerging spatial data infrastructure standards, in an operational environment. Geogratis currently supports Catalogue-based discovery services via the Z39.50 Geo profile, and is expected to provide future online OGC Web mapping and direct-access spatial datawarehouse access services.

The new reprojection and reformatting services provided by Geogratis will also be used to exercise the emerging OGC service specifications within an Intranet environment.

Summary and Readiness Analysis

Organisational readiness

Key organisational issues, related to data access in development of a spatial data infrastructure include:

- Ensuring key government, commercial, and value-added data/related service providers are represented as key stakeholder in the development and implementation of a national spatial data infrastructure
- Collaboration of government data suppliers on coordinated, supportive policies that relate to spatial data access and distribution including: availability of free data, pricing, copyright, and use/integration of electronic commerce
- An access infrastructure and policy that is non threatening to stakeholder mandates
- Support for multiple levels of "buy-in" to the data access infrastructure with a low barrier to entry
- Sustainable long term business models
- Early and clear indication of the role of the private sector
- Early marketing and promotion of the entire spatial data infrastructure program
- Awareness and adoption of international standards

Implementation readiness

Table 1 illustrates the evolution of data access and related spatial data services. Migration from “classic” towards “infrastructure enabled; standards based; and full functioned” is required to bootstrap a national spatial data infrastructure. Both “top-down” and “bottom-up” implementation strategies are suggested. Early adoption, and “best practices” should be followed by key government data providers.

	Classic	Move to online	Infrastructure enabled; Standards-based; Full functioned
Metadata	Ad hoc	FGDC - based	ISO TC211 - based
Metadata Catalogue	Offline, hardcopy Compact disk	Database enabled; Web accessible	Semantic interoperability via search/retrieval protocol (current: Z39.50 GEO; (future: OGC catalogue)
Visualisation	Offline: fax, hardcopy, Compact disk	Web – accessible, Map enabled	Semantic interoperability via OGC Webmapping
Ordering	Phone, fax	E-mail	Web-based, integrated with e-commerce payment
Product selection	Predefined products	Geographic and layer-based subsetting of predefined products	Selection of arbitrary features, layers and feature collections from seamless data warehouse
Delivery	Offline: hardcopy	Offline: softgood (e.g. Compact disk)	Online: File-based for network download (note: file may be generated dynamically from seamless data warehouse) network access service
Packaging/formatting	Offline: hardcopy or softgoods from predefined formats	Online: user specified format selected from pre-generated softgoods	Online: support for user-specified softgood format via dynamic format translation
Payment	Offline: traditional consumer	Online credit-based payment to registered list of consumers	Online e-commerce based, supporting “previously unknown” customers (e.g. online credit-card payment)

Table 1 – evolution of access-related services

References and Linkages

GeoGratis (http://geogratis.cgdi.gc.ca/)

International Organisation for Standards, ISO/TC211 (http://www.statkart.no/isotc211)

Internet Engineering task force (http://www.ietf.org/)

Internet Interoperability Institute (http://132.156.30.81/iii/)

World Wide Web consortium, or W3C (http://www.w3.org/)

Chapter Seven: Other Services

Editor: Jeff DelaBeaujardiere, NASA

Note to Readers: As additional services are built on existing infrastructures, this chapter will include greater depth on the following:
Definition of services (e.g. build on catalogue service, online mapping, data access)
Types of additional services that may exist in SDIs: (transformation, classification, authentication, GIS analysis, data fusion, custom symbolisation, collaboration, gazetteer, referral systems, knowledge base, project and experts directory, applications, algorithms, software directories)
What are the organisational issues of implementing additional services?
What existing software services may be present in your SDI?
What standards may exist for supporting services in SDI?

Context and Rationale

The preceding chapters have discussed three types of services that are fundamental to any Spatial Data Infrastructure: data catalogues, online mapping, and access. A broad range of other geospatial services may exist in SDIs. Other services include, but are not limited to, coordinate transformation, classification, data authentication and validation, data analysis, data fusion, custom symbolisation, multi-person collaboration, gazetteers, processing algorithms, and service catalogues allowing discovery of required services.

The *OpenGIS Service Architecture* defines a number of categories of Geospatial Domain Services. The following fall under the rubric of Other Services in the context of this document:

- Feature Generalisation Services - Services that modify the cartographic characteristics of a feature or feature collection by simplifying its visualisation, while maintaining its salient elements – the spatial equivalent of simplification.
- Geospatial Information Extraction Services – Services supporting the extraction of feature and terrain information from remotely sensed and scanned images
- Geospatial Coordinate Transformation Services – Services for converting geospatial coordinates from one reference system to another
- Geospatial Annotation Services – Services to add ancillary information to an image or a feature in a Feature Collection (e.g., by way of a label, a hot link, or an entry of a property for a feature into a database) that augments or provides a more complete description.
- Image Manipulation Services - Services for manipulating images (resizing, changing color and contrast values, applying various filters, manipulating image resolution, etc.) and for conducting mathematical analyses of image characteristics (computing image histograms, convolutions, etc.).

- Feature Manipulation Services – Services that support creation, quality control methods, analysis, and display of feature collections of interest to an end user.
- Image Exploitation Services – Services required to support the photogrammetric analysis of remotely sensed and scanned imagery and the generation of reports and other products based on the results of the analysis.
- Geospatial Analysis Services – exploits information available in a Feature or Feature Collection to derive application-oriented quantitative results that are not available from the raw data itself.
- Image Geometry Model Services – Support using mathematical models of image geometries, that relate image positions to corresponding real-world (e.g., ground) positions.
- Geospatial Symbol Management Services – Services for management of symbol libraries.
- Image Synthesis Services – Services for creating or transforming images using computer-based spatial models, perspective transformations, and manipulations of image characteristics to improve visibility, sharpen resolution, and/or reduce the effects of cloud cover or haze.
- Image Understanding Services – Services that provide automated image change detection, registered image differencing, significance-of-difference analysis and display, and area-based and model-based differencing.

Omitted from the foregoing list are those service categories covered in earlier chapters: Geospatial Domain Access Services, Image Map Generation Services, and Geospatial Display Services.

While specific GIS software packages may offer one or more of the services discussed here in a proprietary fashion, there are few existing standards and protocols for providing geospatial domain services in an interoperable manner. The *OpenGIS Service Architecture* defines (See Chapter 2) what specific services are included in each service category, but that abstract specification provides no implementation details. Thus, this chapter is simply a placeholder for future implementation advice to be included when available.

Organisational Approach

Implementation Approach

References and Linkages

[1] Kottman, Cliff (ed.), *OpenGIS Abstract Spec topic 12: OpenGIS Service Architecture*, version 4 (1999): http://www.opengis.org/techno/specs.htm.

[2] Buehler, Kurt and McKee, Lance (eds.), *OpenGIS Guide*, 3rd edition draft (1999): http://www.opengis.org/techno/guide.htm.

Chapter Eight: Outreach and Capacity Building

Editor: Uta Wehn de Montalvo, UK

Introduction

This chapter describes the 'softer' elements of a Spatial Data Infrastructure (SDI), focussing on the outreach and capacity building activities that accompany the more technical elements of building a SDI dealt with in previous chapters. Nevertheless, the aspects of implementing a SDI discussed here often present considerable challenges because they depend on the willingness of people in different organisations and institutions to co-operate.

The chapter considers when it makes sense to develop a Spatial Data Infrastructure, how this relates to regional efforts and the Global Spatial Data Infrastructure and how outreach and capacity building activities can be used in the implementation of a SDI.

Contributions from both developed and developing countries have been drawn upon. These are placed along different ends of the spectrum of SDI development; some of these countries have gained much experience in implementing a SDI while others are just beginning.

Several people have contributed with their input or comments to this chapter. Thanks go to Mark Reichardt, FGDC, United States; Liz Gavin, NSIF, South Africa; Camille A.J. van der Harten, SADC Regional Remote Sensing Unit, Zimbabwe; Rita Nicolau, CNIG, Portugal; Bob Ryerson, Kim Geomatics Corporation, Canada; Terry Fisher, CEONET, Canada; Ian Masser, EUROGI; Hiroshi Murakami, Ministry of Construction, Japan; and Steve Blake, AUSLIG, Australia. Acknowledgements go also to the Program on Environment Information Systems in Sub-Saharan Africa (EIS-SSA) for making available the best practice reports on environmental information systems for several countries.

Context and Rationale

When does a Spatial Data Infrastructure make sense?

The continued advances in remote sensing, mapping and geospatial technologies, including an increasing variety of data acquisition capabilities and low cost and more powerful computing capacity, coupled with the development of geographic information system technology, have enabled and increased the demand for geographic information. As the importance of geographic information in addressing complex social, environmental, and economic issues facing communities around the globe is growing, the establishment of a Spatial Data Infrastructure to support the sharing and use of this data locally, nationally and transnationally makes increasing sense.

Without a coherent and consistent SDI in place, there are inefficiencies and lost opportunities in the use of geographic information to solve problems. Furthermore, as spatial technologies are increasingly being used by diverse organisations in developed and developing countries, a number of obstacles add up to a geographic information bottleneck (see Example 1). Lack of institutional co-ordination, insufficient flow of information, overlapping of initiatives, duplication of field activities and results, poor management of resources and insufficient qualification of the technical staff are some of the most pressing problems. In addition, there is a lack of standardised metadata and poor documentation on who is doing what and the types of available information. This has a double negative effective. On one hand, potential data and information users have difficulty finding or having access to needed relevant information and, on the other hand, information suppliers do not know what they have which in turn prevents better organisation of information for dissemination and enhanced value of the information.

It is important to take into account that the longer the harmonisation of stand-alone databases is post-poned, the more difficult it will be to make them interoperable. Costs for integrating stand- alone systems into a SDI concept are increasing exponentially with time and the number of data sets. This suggests that a co-ordinated initiative based on SDI principles should be considered as soon as possible. A feasibility study carried out in Malaysia prior to the implementation of a national SDI concluded that a SDI would present an opportunity with dynamic benefits that would grow over time, culminating in accelerated socio-economic development the nation combined with a reduction in delays in the implementation of projects (http://www.nalis.gov.my/laman/kertas6e.htm).

However, the development of a SDI will rely heavily upon opportunities provided by the socio-political stability and the legal context of a country as well as other important institutional set-ups that might become instrumental while installing a dynamic process of information creation and exchange (see Example 1).

Example 1

Summary of Current Conditions in Developing Countries Around the Globe

Awareness of the value of geographic information and applications is growing quickly, in the public and private sectors.
Growing awareness of the potential of GIS among public sector institutions, non-governmental organisations as well as the private sector means that the use of geographic information systems is increasing every year. However, often the existing spatial data systems are not technically linked and institutional co-ordination is still weak. Most GIS developments started with the implementation of an information component for specific projects. Systems are not designed to ensure smooth data sharing but primarily to respond to specific needs of the host organisation. Although this has helped to design systems with a demand driven approach, this evolution did not create a favourable context for straightforward data exchange.

Co-operation and co-ordination between public sector organisations is limited.
Due to the lack of co-ordination, the different data structures will not be compatible to facilitate data exchange. Although networking relationships exist between people, these are based on individual contacts and are not reflected in an operational co-ordination of activities. There is usually no nation-wide SDI and usually no lead agency has started activities to create one. Many of the systems are still in their installation phase. Where there is metadata at all, different agencies maintain it using different formats and tools. More generally there is a lack of common elements that could facilitate data exchange such as same working scales, same GIS software, and the completion of a national database which could be used as standard basic information layers. In many instances, there is no copyright law and most public agencies need to market their product in order to find additional resources to maintain and update their data. Only very few institutions have already started to define clear data exchange policies to disseminate their information.

Development and implementation are very internal, stove-piped and do not favour data sharing collaboration.
The spatial databases being built up are "stand alone systems", using individual philosophies and technologies (concepts, structures, hard and/or software). Most of these implementations are technology and/or donor driven and as such are isolated implementations, insular databases under construction, and related to specific environmental issues. The whole problem is exacerbated in developing countries since different agencies are often supported by a different donor. Each donor tends to encourage its own solution – often resulting in interagency competition instead of co-operation. Few of them are ready to deliver some outputs, none of them are fully operational yet. Communication between the different implementations is usually not possible technically because common communication standards for data exchange are missing. The exchange of information between institutions and teams ranges from limited to non-existent. Often the relation between the implementations is characterised by competition rather than co-operation. Existing systems primarily serve the purpose and mandate of their host organisations, who are only now beginning to co-operate and co-ordinate. There is very limited co-operation and co-ordination between public sector organisations.

Most of the motivation to employ geographic information and tools is still internal to institutions to serve their primary needs. Outreach and education are not being emphasised.
The majority of the institutions are motivated by their own mission and therefore to a great extent do not subscribe to national policy objectives. Existing systems serve primarily their own clientele, without concern for the needs of other potential users. This leads to the duplication of efforts and sometimes inefficient use of resources, both financial and human. Sharing information in a fully transparent manner is not the main characteristic of the usual communication culture. Communication is instead linked to hierarchy and authority. Since the success of a SDI is based to a large extent on cross-sectoral networking and access to information, the inherent organisational "communication culture" impedes the build- up of an efficient SDI.

There are few national policy initiatives underway to encourage sharing and collaboration on geographic data and practices.
There are only a few formalised institutional links to share data. Practically every organisation has its own way of producing digital data. Some departments are developing their own data standards including classification schemes for their own use. The awareness of copyright issues is rising, but there is often a complete lack of policy around information management - it has not been addressed simply because it is not seen as a priority.

Vertical organisation within government and administration is limiting cross sectoral communication.
Due to the strong vertical organisation culture of government and administration, there is no real encouragement of cross sectoral communication. Each ministry or department undertakes its own mandate, trying to create it's own database and information system, following it's own needs, point of views and priorities. Information is handled in a strictly vertical direction, following hierarchies. Information seems always to be linked to persons and their status within the hierarchy. Cross sectoral information exchange is strictly limited to informal organisation. The handling of information is a political issue, a cultural topic.

Access to information is hindered by a lack of transparency
Access to information is not only a question of ownership and attitude to communication. Transparency is not yet the main characteristic of communication culture and remains a major problem. Nobody really knows who disposes of what, where what is available or who is in charge to produce what. Without an overall information concept, without clear mandates, tasks and responsibilities, without a metadata-database, access to information remains a casual event, a question of personal relationships and good or bad luck. Users of information have to know about and in some cases hunt for information. To collect precise information one needs either a very good personal network, based on personal relationships, or a lot of time and good nerves. The major technical obstacles to data sharing reside in the lack of application of a national standard for spatial data, incompatible classification schemes and the almost total absence of data documentation or metadata. Additional difficulty stems from restrictions on spatial dissemination for maps of border areas.

These problems are not exclusive to developing countries. A fundamental problem underlying data sharing and distribution is the belief that one gains power and influence from withholding information and controlling it. In fact, true power is held by those who distribute the information and whose information is used by senior political levels. Once this leap of faith is taken, as it has been in several countries, data sharing becomes remarkably easy.

Example 2

The national SDI in the US: Much of what is today's U.S. Federal Geographic Data Committee (FGDC) and the National Spatial Data Infrastructure (NSDI) have roots in the concern by Presidential Administrations since the 1950's to better co-ordinate the operations of agencies engaged in surveying, mapping and related GIS functions across government. Two major activities to drive co-ordination were the Office of Management and Budget published Circular A-16 in the late 1950's, and the activities of a federal mapping task force convened in the early 1970's. The Task Force was charged with studying the possibility of consolidating geographic information (GI) functions across the federal government to reduce potential duplication and overlap, and to potentially reduce costs. Pressures to consolidate Government GI functions continued and in the early 90's the US Government recognised the need to establish a sustaining spatial data infrastructure as part of its National Information Infrastructure. With the advancement of technology and the increase in the personal computers, there was an accelerated explosion of digital information production from a multitude of federal, state, local, other public and private sources. The need for a compatible infrastructure to find, share, and exploit information across jurisdictions became a common goal of many organisations to reduce duplication and improve support to users, and better co-ordinate the operations of agencies engaged in surveying, mapping and related GIS functions. The FGDC was created in 1990. The Committee was created to "promote the co-ordinated development, use, sharing, and dissemination of geographic data". Specific support was requested from several key federal agencies involved with geospatial missions. Today, the FGDC has added more key federal departments, agencies and others will soon become a member as well. The role of other Federal Agencies is expanding as they realise the spatial significance of their social, environmental, economic data, and the FGDC focus now is moving toward getting these data types (such as crime and health data) recognised as national spatial data infrastructure components. The FGDC has also expanded its partnerships to include state, local, tribal governments, and representatives from the GIS industry and academia.

The national SDI in Australia: In Australia, the initial impetus came from the Australia New Zealand Land Information Council (ANZLIC), the peak inter-governmental body for spatial data issues. Each State and Territory and the Commonwealth were represented but there were no industry stakeholders. Some 3 years of the ASDI was spent scoping the size of the tasks ahead and allocating jobs and lead agency status for specific tasks. The recent 12 months have seen the operationalisation of the SDI programs in each of the States and Territories.

Survey of national and regional SDI's: A global survey of many national and regional SDI's can be found at http://www.spatial.maine.edu/harlan/GSDI.html gathering baseline information on the nature and characteristics of the national SDI's that are currently being developed. For each national or regional entry, the following information is provided:

the type of organisation(s) taking the leadership in the co-ordination and development of the SDI,
the types, categories or forms of spatial digital data made available through the SDI,
the technical and organisational access mechanisms of the SDI
private sector involvement in the SDI
public domain data sets

legal mandate or formal orders behind the establishment of the SDI
the components of the SDI
most pressing challenges.

Another important resource considering different SDI development strategies can be found at http://www.gsdi.org/canberra/masser.html More infrastructure developments are provided at http://www.gsdi.org/

These sources suggest that the concepts of core data (or framework data), data standards, clearinghouses and metadata are well accepted as parts of SDI's in many nations around the world. From the standpoint of global SDI development, these are areas where we collectively should place our near term efforts in gaining international agreement where possible.

A SDI makes sense at the local, national, regional and global level where the overlap and duplication in the production of geographic information is paralleled by insufficient flows of geographic information among different stakeholders due to a lack of standardisation and harmonisation of spatial data bases. Once the importance of providing geographic information as an infrastructure similar to road and telecommunication networks is recognised, it makes sense to ensure that a consistent Spatial Data Infrastructure at the local, national and global level is developed.

The 'ideal' SDI: The characteristics of what may be described as an 'ideal' SDI are outlined below;

- There is a common spatial data foundation organised according to widely accepted layers and scales (or resolution) that is available for the entire area of geographic coverage (parcel, neighbourhood, city, county, state, nation, etc.) to which other geospatial data can be easily referenced.
- The foundation (or core) data is readily accessible and available at no or little cost from user-friendly and seamless sources to meet public needs and encourage conformance with it by producers of other geospatial data.
- Both foundation and other geospatial data, as required and specified co-operatively by data producers and users, is updated according to commonly accepted standards and measures of quality.
- Thematic and tabular data are also available on terms not incompatible with the foundation data.
- Cost-effective, geospatial data produced by one organisation, political jurisdiction, or nation is compatible with similar data produced by other organisations, political jurisdictions or nations.
- Geospatial data can be integrated with many other kinds or sets of data to produce information useful for decision makers and the public, when appropriate.
- Responsibility for generating, maintaining, and distributing the data is widely shared by different levels of government and the private sector. Governments take advantage of private-sector capabilities available at reasonable prices rather than maintaining dedicated capabilities.

- The costs of generating, maintaining, and distributing such data are justified in terms of public benefits and/or private gains; overlap and duplication among participating organisations is avoided wherever possible.

(United States National Academy of Public Administration 1998)

Organisational Approach

Principles of the GSDI

At the 2nd GSDI Conference in 1997 the Global Spatial Data Infrastructure (GSDI) was defined as *".. the policies, organisational remits, data, technologies, standards, delivery mechanisms, and financial and human resources necessary to ensure that those working at the global and regional scale are not impeded in meeting their objectives."*

The GSDI is intended to be non-competitive, collaborative, and to build on and unify common activities in the field of geographic information exchanges and harmonisation. The GSDI is envisaged to support trans-national or global access to geographic information and it is seen by many as central to the response to the challenge of global sustainable development. It is an effective promotion of national and regional Spatial Data Infrastructures.

Examples of how these principles are promoted and implemented at the regional and international level are given below.

Example 3

Regional collaboration: The European Umbrella Organisation for Geographic Information (EUROGI) was set up to foster geographic information outreach and capacity building at the regional level. EUROGI's objectives are to support the definition and implementation of a European geographic information (GI) policy and facilitate the development of the European Geographic Information Infrastructure (EGII). It also represents the European view in the development of the Global Spatial Data Infrastructure (GSDI) and is the European regional contact for GSDI. In a more general sense EUROGI tries to encourage the greater use of GI in Europe through improved availability of and access to GI, the removal of legal and economic constraints to use, and the promotion of the use of standards. As an association of associations, EUROGI works towards the development of strong national GI organisations in all the European countries with particular emphasis on organisations within the Central and Eastern European countries.

International collaboration: The United States has been a recognised world leader in the development and use of geographic information and related technologies. Recently, on behalf of the organising committee of a conference on Global Spatial Data Infrastructures, the FGDC conducted a survey of spatial data infrastructure activities around the world. This survey showed that there are a growing number of nations, which are either developing or planning to develop spatial data infrastructures. These initiatives, while reflecting the specific needs of the various nations, were found to have many components in common with each other. These same components are also part of the United States National Spatial Data Infrastructure, which is becoming a model that is frequently looked to and used by other nations as they consider ways in which they can better co-ordinate and use geographic information. The FGDC is increasing its focus on the international and global community to help assure that NSDI development is accomplished so that data, practices, and applications can be shared wherever possible to address transnational, regional, and global economic, environmental, and social issues. The FGDC is an active supporter of the GSDI, it is pursuing nation to nation agreements to foster SDI collaboration on topics of mutual interest, and it is a strong proponent of the formation of a Permanent Committee of the Americas to address the infrastructure issues specific to the nations in the Americas.

Different levels of international collaboration: GeoConnections, the program responsible for implementing the Canadian Geospatial Data Infrastructure (CGDI), believes that international partnerships are important at many levels. For example, the Canadian clearinghouse is interoperable with the US and Australian clearinghouses and the Canadian program has supported the development of access tools that are being reused in the US and Canada. Canadians have been very active in many of the international standards activities and now, as the infrastructures are being implemented, there is a significant opportunity to co-operate with international partners and industry in the development of implementation specifications, such as the Open GIS Consortium Catalogue Services and Web Mapping Testbed.

Realisation of the GSDI

The stakeholders and interested parties in the development of the GSDI were identified at the 3rd GSDI Conference (1998) in Canberra, Australia:

"The achievement of GSDI will depend upon partnerships among many groups including industry, consumers, academia and government. GSDI must develop outreach activities to ensure that institutions and organisations that can and will benefit from an improved global spatial data infrastructure have an opportunity to participate. At this meeting it was obvious that national mapping organisations/agencies, state level mapping organisations/agencies, industry, academia and a variety of governmental agencies are very interested in GSDI development.

- ***National mapping organisations/agencies***

National mapping organisations/agencies play a key role in ensuring that accurate, up-to-date geospatial framework data are developed and maintained. Such data are key to, among others, the promotion of sustainable economic development, improvement of environmental quality, resource management, upgrading public health and safety, modernisation of governments either local, national or regional, *and the responses to natural and other disasters. Therefore such organisations play a vital role in facilitating the development of a GSDI.*

- ***Industry***

Industry is working to provide technology, data and services in support of GSDI activities. In particular, industry plays a key role in ensuring that effective information technologies (consistent with standards and specifications being developed by such groups as ISO and OGC) exist and that these technologies support GSDI requirements. Therefore it is imperative that such organisations play an important, proactive role in the development of a GSDI.

- ***Other agencies, organisations and institutions***

There are many other agencies, organisations and institutions that collect and use geospatial data that along with national mapping organisations/agencies and industry can and should play an important role in GSDI activities. It is important here that ways be sought to encourage cooperation, collaboration and communication among as many GSDI stakeholders as possible.

- ***National and regional SDI initiatives***

There are a growing number of significant SDI initiatives at national and regional level that can and will act as a stimulus to GSDI development. Several of these initiatives were highlighted at the 3rd GSDI Conference - national developments in countries such as Malaysia, Hungary, Australia, New Zealand, USA, UK, Canada - regional developments in areas such as South America, the Baltic Sea Region, Europe, Asia and the Pacific. These initiatives are now being documented in several ways and this documentation provides a valuable resource for proponents of the GSDI."

The GSDI acts as an umbrella organisation that brings together national and regional committees and other relevant international institutions. As such, it provides an opportunity for pro-active countries in SDI implementation to be generous with their ideas, knowledge and experience from implementing SDI at various levels. Rather than imposing a regional or national SDI overnight, tangible projects such as the SDI Cookbook provide an opportunity to assist other countries in the development of a SDI. It can be considered a pool of resources that different countries or regions can tap into and contribute to.

Example 4

Pooling resources: The Global Mapping initiative, Globalmap, promoted by the Geographical Survey Institute of Japan, is a key pool of resources for GSDI development to exchange institutional and technological experiences and standards among many countries. The US FGDC, in collaboration with other nations, has helped to seed many common standards and best practices. Japan has adopted its National Spatial Data Infrastructure Promoting Association (NSDIPA) as a reflection of the US NSDI. Other nations have adopted or have based their NSDI's on FGDC practices, standards, and framework concepts. Some of the ISO TC 211 standards are based on FGDC developed standards (for example, Metadata). Globalmap exemplifies a global “framework”, ISO TC211 the reference standards environment needed to assure data sharing between jurisdictions.

It is not necessary to implement a national SDI before approaching a regional SDI. Special attention should also be given to regional and international co-ordination and co-operation with other countries and with international institutions and donors. A joint approach to SDI within a particular region, for example, would not only save a lot of energy and expenses. The potential for synergy would also be considerable, since it would be possible to enable cross-border exchange of data and information and supporting infrastructure elements such clearinghouse software and metadata structures.

Standards and models for a common SDI do not have to be reinvented by each country. A common vision and common standards throughout Southern Africa, for example, would improve the efficiency of national and regional SDIs. This would entail effective exchange of experiences and results, a co-ordination and division of work within existing national institutions in the region, including NGOs and representatives of the donors involved, an efficient partnership with a non permanent joint steering committee as a co-ordinating body.

Implementation Approach

How does one build a successful SDI as part of GSDI?

Many success stories can be reported that are encouraging to those just starting out on SDI development. However, it may be equally helpful to know that they are not alone in encountering difficulties. It may take some time until efforts bear fruit and different strategies and approaches may need to be considered to get people on board (see Example 5).

Example 5

Delays in success: As the GI community in South Africa has frequently requested, the technology for capturing and publishing metadata has been put in place by the National Spatial Information Framework (NSIF) directorate in charge of implementing the national SDI. For the users, there are no costs associated with this clearinghouse (the Spatial Data Discovery Facility). However, despite the best efforts by the NSIF, the fact that the clearinghouse is available does not seem to be in people's heads yet and they still come out with statements like "what we really need is ...". Moreover, people do not contribute metadata to be included in the system.

Yet this lack of awareness and participation is likely to be temporary. In a recent survey of the South African GI community, about 70% of the participating organisations considered the clearinghouse provided by the NSIF a very important facility but only a small percentage indicated that they possessed the necessary metadata skills (Wehn de Montalvo 1999). Once these skills are in place, the use of, and contributions of metadata to, the Spatial Data Discovery Facility are likely to increase.

While there is no prescriptive recipe for building a SDI, the following aspects have emerged as 'lessons learned' from the international arena of SDI developments. They may need to be adapted to the specific political system and social context within which a SDI is being developed.

- **Build a consensus process: build on common interests and create a common vision**
- **Clarify the scope and status of the SDI**
- **Exchange best practices locally, regionally and globally**
- **Consider the role of management in capacity development**
- **Consider funding and donor involvement**
- **Establish broad and pervasive partnerships across private and public sectors**
- **Develop clearinghouses and use open international standards for data and technology**

Creating a Common Vision: A common vision can be an extremely powerful management tool, especially in complex projects, where multiple parties have to co-operate in order to reach a consensus. A vision of the future nation wide SDI could help to streamline future activities towards a mutual objective. A mutual objective can open perspectives and offer security in periods of change.

Even in contexts where the community of technicians involved in GIS development is small enough to allow all the members to know each other, there often is no apparent willingness at the institutional level to co-ordinate and harmonise the development of the systems. The development of a SDI will require cultural and organisational changes so as to manage the whole shifting process. This entails mobilising resources so that people in different organisations can adjust.

Example 6

Creating a common vision: The Australian experience in establishing a national SDI shows that getting people on board has been a long process and has been driven by ANZLIC in terms of awareness raising and making the major components of the ASDI more tangible. Informal collaboration is fairly smooth. As the number of Australia's spatial data stakeholders is quite small, most people know one another, so ideas and knowledge get exchanged quite easily. Formally, ANZLIC is the formal process to endorse collaboration activities, but in reality people just go to the individuals or agencies who have worked in specialist areas to get advice and help. The ASDI is therefore not too regimented. The States, Territories and the Commonwealth are all working together on most national implementation projects such as the Australian Spatial Data Directory (ASDD), Australia's fully distributed metadata directory.

Masser (1999) has summarised the objectives of most national SDI's. These are intending to promote economic development, to stimulate better government and to foster environmental sustainability. A selection of SDI vision statements is provided below.

Example 7

Selected vision statements of SDI initiatives:

Colombia (ICDE):	*http://www.igac.gov.co/indice.html*
Europe (EUROGI):	*http://www.eurogi.org/objectives/*
Finland (NGII)	*http://www.nls.fi/ptk/infrastructure/vision.html*
United Kingdom (NGDF):	*http://www.ngdf.org.uk/*
United States (NSDI):	*http://www.fgdc.gov/nsdi/strategy/goals.html*

But a common vision for a SDI may be missing or hindered by reasons such as culturally based resistance. In many instances, information is linked to personal power and tends to be strictly controlled in a top-down manner. This "personalised" approach to information may be one important reason for a lack of a shared SDI approach and also hindering the various stakeholders to produce a shared common vision of a national SDI. High-level commitment and support may be crucial in implementing a change in culturally-bound attitudes.

A common and shared vision about spatial data collaboration and co-operation may fundamentally change the landscape for a nation wide exchange of data and information. In order to get the various stakeholders on board, it may be essential to insist on joint development of a common vision. This may entail a cultural change in the attitude towards information and the exchange of information, a new approach how to manage and share information. The process of getting the concerned parties involved to accept and to actively support the idea of a SDI will need both a strong lead and a lot of creativity in order to minimise unnecessary resistance and not to de-motivate or suffocate creative initiatives.

The vision needs to be developed jointly and shared with the conceivable stakeholders and indicate the incentives for developing a SDI so that people are mobilised to change their behaviour in accordance with the shared vision.

A participative approach to co-operation and co-ordination should be considered in order to build on common interests. This also entails initiating a participative process among the representatives of the already existing database systems. It would make sense to bring the up to now independent system owners, stakeholders, donors, representatives of international organisations active in the field of GIS, soft- and hardware suppliers, and database managers, including their technical staff, to sit together at a round table in order to develop a common concept of a nation wide SDI.

The common standards and procedures the stakeholders will have to agree on will not necessarily fit into their actual database set up but a participative approach and a transparent decision making process will help them to understand the basic questions and to accept the resulting needs for change. Participative processes and transparent decision making are strong arguments to motivate the independent parties to invest their resources in a common project.

The vision needs to be communicated widely using various media to reach all stakeholders. Plans should be developed and implemented regarding the dissemination of information on SDI activities that are under way, including the information about the SDI components, available technological best practices, and the promotion of the use of existing technologies and standards to support the development of a SDI, for example by establishing WWW pages on the Internet or using printed media or CD-ROM where Internet connections are limited.

SDI Scope and Status Clarification: Two broad categories with respect to the status of a national SDI can be identified (Masser 1999), i.e. a SDI resulting from a formal mandate (as was the case in the US, for example) and a SDI growing out of existing spatial data co-ordination activities (as was the case in Australia). While a formal mandate benefits from the provision of funds, existing co-ordination activities provide a base for collaboration. The scope of a SDI may be all-inclusive or focusing on a subset of stakeholders, such as public sector, private sector, or NGOs, with voluntary or mandatory participation. Regardless of which category a SDI falls into and regardless of the breadth of its scope, both should be clarified as early as possible.

An active co-ordination body (committee or commission) to co-ordinate tasks and provide leadership during the process of creating a national SDI should be considered. This would need to be sufficiently empowered to carry out the co-ordination task. In order to implement a SDI, it may not be necessary to establish new organisations and institutions. Instead, existing ones could be strengthened. This would require a revision of the mandates of that institution to ensure that it is well equipped to deliver.

However, the promotion of an existing institution to the co-ordination body for a SDI needs to be carefully considered. The institution needs to be chosen carefully so as to be aware of potential conflicts of interest that may be perceived between the institution's existing mandate and the additional SDI-related activities. For example, a National Mapping Organisation may end up carrying out the SDI co-ordination task and policy development while also acting as a major data producer. This may hamper the support for the SDI initiative from potential participants that could perceive it as biased. Example 8 demonstrates that although it may take some time for the co-ordinating body to gain support, a crucial element to success is how its mandate is perceived.

Example 8

Perceived mandate: In Portugal, the national SDI (SNIG) is co-ordinated by the National Center for Geographic Information (CNIG). CNIG is not a major data producer, like many agencies in other countries that are responsible for co-ordinating a national SDI. Development of the SNIG was slower than expected mainly due to lack of available digital GI and the incipient computer technologies used by most GI producers. The fact that the CNIG is not a major data producer facilitated the interactions with the GI producers, as they recognised the role of the CNIG as being a complementary one that did not harm their own mission.

The task of promoting and developing a SDI is not restricted to the public sector. In Japan, for example, the private sector is a major driver behind the establishment of a national SDI (see Example 9).

Example 9

Private sector involvement: In 1995, the Government of Japan established a Liaison Committee among Ministries and Agencies on GIS that is to provide SDI-like functions in the Government in implementing a national SDI in Japan. Private companies in Japan set up the National Spatial Data Infrastructure Promoting Association (NSDIPA), a non-profit organisation to promote the concept of national SDI in Japan. The activities of NSDIPA are aimed at gaining wide awareness of the necessity of the National Spatial Data Infrastructure. It is a group that strives towards the benefits of society and fosters a new information services industry by demanding activities of the government, municipalities and other organisations and by sharing this information with both the public and private sectors.

The representatives of all major sectors or interest groups should be involved. The co-ordinating body, once nominated and appropriately mandated, can then produce a series of activities which need to be accomplished with deadlines and output milestones. The implementation process should be approached in a multi-disciplinary and multi-sectoral way. All related organisations will have their role to play in the SDI development process.

Working groups constitute the platforms for more collaboration among stakeholders by pooling resources and harmonising initiatives to avoid duplication. The involvement of stakeholders is a key issue for the future development of a SDI.

Exchanges of Best Practice and Awareness Creation: Lessons in awareness creation about SDI can be drawn upon from various countries. These suggest that presentations and publications are just some of the activities that can be pursued to advocate and advance SDI development. Networks of communication (see Example 10) can also play an important role. A list of activities includes:

Outreach through support for SDI from high-profile individuals.
Promotion of SDI principles through presentations.

- Education through workshops, training courses and material.
- Provide "train-the-trainer" technical workshops to explain the origins, purpose, and strategies for implementation of the SDI-endorsed standards.
- Use pilot projects to demonstrate the value of spatial data and a SDI to improve decision making in communities.

Establish networks of communication to enable participants to exchange experiences with SDI implementation.

Facilitate information sharing through newsletters, web pages, and publications: regularly inform interested parties of SDI-sponsored activities and initiatives.
Provide a forum for debate, analysis and the identification of issues relevant to SDI development.

- Help interested parties or groups to use the spatial data clearinghouse to locate sources of data, training and expertise.
- Offer interested parties the opportunity to participate in Working Groups and Subcommittees, as appropriate.

Example 10
Networks of communication: EUROGI, the European Umbrella Organisation for Geographic Information, seeks to raise awareness of the value of GI and improve the sharing of knowledge between members themselves and between EUROGI and the European Commission. Communication is facilitated through on-line discussion forums and EUROGI directories where people are able to tell others about their activities by completing a form to add information to a directory or search a directory to read about other peoples' activities.

Examples of how demonstration projects can be used to create awareness of the usefulness of a SDI are detailed in Example 11.

Example 11
Community Demonstration Projects: The FGDC has worked with the Administration and Federal Agencies to promote several Community Demonstration Projects (http://www.fgdc.gov/nsdi/docs/cdp.html) across the country. These NSDI based pilots are designed to demonstrate the value of spatial data and the NSDI to improve decision making in communities. The Demonstration Projects address a number of issues including flood management, local/regional crime management, Citizen-based land use analysis, environmental restoration. The NPR and the FGDC jointly queried FGDC membership to seek interested communities, offering only in-kind federal help (federal staff, training, etc but no dollars) to maturate the projects. Shortly after the selections, the six selected communities joined together to apply for a grant under the Government Information Technology Services Board (GITS). They were awarded over $600,000 dollars as part of their grant request. These projects are expected to report back in May 2000, with the detail of each effort utilised to help articulate the value of NSDI to enhance place-based decision-making, and to help communities understand the costs and processes associated with establishing NSDI operations.

Community-based capacity building: In 1998, the FGDC working with OMB and its federal agency representatives began a $40M multi-agency budget initiative to accelerate the application of the NSDI to improve the decision-making power of communities in addressing livability issues. The Community Federal Information Partnership (CFIP) (http://www.fgdc.gov/nsdi/docs/schaeferbrief/index.htm) was first announced by the Vice President as part of a Sustainable Communities Speech delivered in 1998 at the Brookings Institute. The C/FIP would provide grants to communities to activate capability and tools for place-based decision-making, and would provide federal agencies with additional funding to help make their spatial data more available for public access. The outcome of CFIP for fiscal year 2000 is still being worked through the congressional budgeting process at the time of this writing.

Exchange of best practice: The FGDC has developed metadata, Clearinghouse, data Standards training, and has developed and offers metadata tools. Assistance is provided by FGDC and FGDC trained partners to local, state, federal, tribal, and international organisations seeking to establish or improve their SDI.

The Role of Management in Capacity Development: An important barrier to change is an organisation's capacity to adopt new standards and technologies. While the introduction of specialised software, for example for the creation of a geospatial catalogueue, is relatively easy, its effective use depends on the technical capabilities as well as organisational support. Awareness creation of SDI components should be considered down to the lowest level and with strong management support and leadership. Capacity development should be a prime concern of senior management. It includes the theoretical issues and the practical hands-on capabilities to implement the SDI components.

This issue of building local capacity will to be a major constraint to the success of a SDI in many developing countries. As job specific technical competencies will be stipulated, it will be necessary to review positional titles, remuneration packages and salaries. The staff rotation system in the "Department of Geological Surveys" in Zimbabwe is a case of "best practice" in how "brain drain" can be avoided and serves as an example of how staff can be motivated within a "Learning Organisation". This system is designed to enhance the capacity of personnel within the department, therefore reducing the need for external recruitment of technical staff.

The personnel resources for SDI in many countries are very limited since most of the GIS implementations being built up are understaffed. A pool of qualified staff has to be created if the projects are to become sustainable. What makes it difficult for countries such as Zimbabwe, for example, is not only the number of specialists required, but also the working conditions offered. "Brain-drain" is a serious problem: the fact that skilled personnel are leaving their jobs too often, too soon. Human capacity development and long term career planning should be of prime concern to senior management. It includes the training, theoretical issues and practical hands-on capabilities to implement projects and programs, as well as the working conditions. Working conditions need to be considered not only with respect to salary, but even more importantly with respect to the work climate, motivation and professional perspectives.

Example 12
Compensating for high staff turnover: One of the US NSDI Community Demonstration Projects is taking place in the Police Department in Baltimore, Maryland. The Police Department has come to realise that a SDI is good for managing vital crime data in addition to the classic mapping data that many rely on for the base mapping. The Baltimore Police budget is tight, they have a high turnover in staff. By capturing metadata and using clearinghouse capabilities, they can better assure the proper management of critical crime data used by the

department and throughout the region as part of a regional crime management collaboration between community police organisations.

Senior management of all concerned institutions should consider the development of standards a priority. They should closely supervise technical work groups and assure that the desired results will be produced. Matters like the standardisation of data and the harmonisation of classification schemes cannot be left to technicians alone because they entail political decisions. Senior management should be acknowledged as a driving force behind the build-up of a SDI.

Funding and Donor Involvement: Funding and adequate resources can present a major constraint to SDI development when awareness of the importance of SDI is lacking at the local, national or regional level and there is no existing SDI-like initiative or a mandate to develop a SDI to which sufficient funds have been assigned.

Nevertheless, in order to ensure funding, it may more persuasive to potential funders to have something to show already (for example, a clearinghouse system) rather than a concept document alone. This does not have to involve huge costs since clearinghouse components are available free over the Internet (**link to Chapter 4**). In addition, justification for the limited cost of this initial development may well be found within existing projects or initiatives (for example, documenting data holdings is a part of sound information management).

Innovative use of resources can ensure that funds stretch a long way. For example, with a 'carrot and stick approach', incentives can be created for the adoption of SDI principles. Using small, non-repeating grants to stimulate the development of the application layer of the SDI can work well where there is broad base of existing expertise that can be encouraged (see Example 13).

Example 13
Program of Grants: The FGDC in the US has maintained a relatively small but persistent Cooperative Agreement Program of Grants (CAP) to help communities validate and initiate NSDI concepts (http://www.fgdc.gov/publications/publications.html). The FGDC initiated the CAP program to provide seed money to stimulate co-operative activities among organisations to begin implementing the NSDI. Rooted in the premise that building the NSDI is a shared responsibility and collaborative efforts are essential for its success, the CAP program has worked to seed 270 NSDI resource sharing projects across the county involving more than 1300 organisations. These projects have resulted in helping state governments, libraries, universities, local government organisations, and private sector entities to become stable contributing sources on the NSDI. While the FGDC funding level for the CAP has been somewhat limited ($1 - $2M yearly), annual funding has been persistent since 1994, and recently the number of grants awarded has been increasing – communities appear to be doing more with less.

Reports of different SDI funding mechanisms from Australia and Portugal suggest that the provision of central funds is an important contributor to accelerated SDI development (see Example 14 and 15).

Example 14
Decentralised funding: In Australia, there is no major national funding allocation for the ASDI (unlike the US and Canada). Each jurisdiction (States, Territories and Commonwealth) are all financing their own programs. The States and Territories in Australia are each developing their SDI's, so in fact the ASDI is all the individual jurisdictional SDI's puzzled together. This approach has some drawbacks. It would be more coherent if a national SDI pool of funds was available to draw on and bring influence to bear. Industry groups are not really engaged in Australia yet to the same degree as in the US or Canada. Making the ASDI politically attractive for large scale national funding has not occurred yet but continuing efforts are under way. One notable success was the establishment of the Australian WWW Mapping Consortium as a full member of the OpenGIS Consortium (OGC). 23 Australian industry, R&D and Govt Groups have all come together to share ideas and work on an "Australian Web Mapping Testbed" which making good progress, allowing them to provide some new input into the OGC process.

Example 15
Centralised funding: The creation of the Portuguese national SDI, the SNIG, was endorsed by public funds. The approval by the Portuguese Government and by the European Commission (at the end of 1994) of a program integrated in the Regional Development Plan 1994-1999 did cover a specific budget assigned to the support of the SNIG development. Part of the funding was used to speed up the creation of digital geographic information, namely for conversions of existent geographic information into digital formats, and for the purchase of satellite data and existent digital topographical data for local GIS implementation in municipalities. Another part of these funds was used to provide major public data producers with Internet servers, routers and communication infrastructures. A small fraction of the funds is still being used to build WWW interfaces and applications to facilitate the access to geographic information available at the different institutions integrated in the SNIG network. In the Portuguese case, funding was a major factor that helped the fast development of SNIG since 1995. It did accelerate a process that would have taken years to grow up. At the present moment, a total of 117 public institutions, including almost all GI producers, have joined the Portuguese spatial data infrastructure.

GIS implementations in developing countries are often functioning under special conditions that need to be considered during the initiation of a SDI at national or regional level. In many countries the lack of local financial resources means that GIS implementations are not financially sustainable and therefore depend primarily on donor funds. Usually donor support for these projects is provided under certain conditions such as a time limit for implementation after which there are no further disbursements of funds. The future of many of these systems is uncertain beyond the end of international assistance.

Another aspect of donor-funded GIS implementations is that often the projects have been initiated by donors according to their own objectives and little attention has been paid to the requirements and capacities of the host organisations. The result is that there is insufficient co-ordination of the technical support and funding activities of different donors. In some cases donors may not be willing to work with each other and this can impose limits on the co-operation or data exchange between projects that are funded by different donors. A lack in capacity to co-ordinate donor activities coupled with competition among the donors themselves can hamper a SDI initiative.

Under these conditions, the co-operation with donors is a critical aspect of the development of a national SDI. While the existing co-operation should not be exposed to strain, a co-ordinated SDI-based approach would change the priorities for GIS implementations. This potential conflict could be avoided if donors would be invited as partners to take part in the participative process defining the components of a nation wide SDI.

In order to develop (or renew) a national SDI in a (multi) donor-funded GIS context, a useful approach has been developed by Ryerson and Batterham (2000). This approach entails an evaluation of GIS projects with regard to:

clients' needs and desires,
an assessment of the recipient country's capabilities in terms of meeting those needs,
an assessment of related activities by other donors,
an assessment of current technology and its direction,
donor country capabilities and capacity if the aid is tied, and
costs.

The issue of building up local capacity will continue to be a major constraint to the success of a SDI in many countries. Long term projects require not only long term financing but also long term planning in the field of human resource capacity building. What requires to be worked on is the issue of sustainability of the initiatives with respect to capability to keep up with the technology shifts and capacity of local personnel. The build-up of a GIS implementation is a long term investment, taking many years until return on investment is visible. Therefore, the ever scarce budget resources are likely to be invested in more urgent projects with prospects of short term successes and returns. This means that participants of such a SDI will remain dependent on donor funds for quite some time.

Example 16
Initially funded by Donors, the SADC Regional Remote Sensing Unit (RRSU) in Harare, Zimbabwe, has been integrated in the organisational structure of the Southern African Development Community (SADC) since 1998. The Unit is funded by the 14 Member-states (Angola, Botswana, Democratic Republic of Congo, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Secheylles, South Africa, Swaziland, Tanzania, Zambia, and Zimbabwe), and still receives some additional donor contributions.

The SDI work implemented by the RRSU was never part of the original work plan. It was identified when GIS technology was to be used for basic analytical procedures. This could not be done because the data sets were incomplete or incompatible. At the time the SDI work started, the Unit was still depending on donor assistance and technical assistance from the FAO. Consequently, changes had to be made to the work programme, something which had to be discussed with the donor and technical assistance partner.

From the regional and international partners in the development (the data suppliers) of the RRSU spatial data sets, no financial contributions were required.

The RRSU spatial data sets were originally developed for GIS-based applications in support of early warning for food security. However, the data sets are being recognised as one of the major spatial data base developments in the SADC region, and for this reason, the RRSU continues to attract donor funding. The spatial data base activities were originally not foreseen as a major task - but this has changed considerably over the years. (http://www.zimbabwe.net/sadc-fanr/intro.htm)

Broad and pervasive partnerships across public and private sectors: Co-operation and partnerships across different levels of the public sector and with the private sector can be helpful at every stage of SDI development to collect, build, share, and maintain spatial data.

Since no one organisation can build a SDI, collaborative efforts are essential for its success. The FGDC in the US encourages federal, state, local, and tribal governments, academia, the private sector, and non-profit organisations to work together within a geographic area to make geospatial data available to all. So-called 'cooperation groups' are formed that enable all parties to participate in, and contribute to, the national SDI in the areas of their strength and expertise. Guiding policies and procedures for these cooperation groups have been developed (http://www.fgdc.org/funding.html). Co-operation among Federal, State, local, private, and academic sectors is expected to be based on shared responsibilities, shared commitment, shared benefits, and shared control aiming at improving the spatial data delivery system (see also Example 17).

Example 17

The task of building relationships to further the implementation of the NSDI in the US has been a large and continuing effort. The effort has been made difficult by the fact that organisations, functions, and responsibilities are diverse and spread out across the country. Initial efforts concentrated on FGDC initiatives to build relationships with Coordinating Groups that have formed to represent issues within the States, and with Organisations and Associations that represent levels of government of key interest groups nationally. This has helped focus the work of the different groups and has established strong linkages with some of the key elements needed for a long-term national network of partnerships. The FGDC's efforts has also been helped by the fact that many in the United States recognise the value of geographic information to the decision making needs of communities. Geographic information is collected at all levels. Most of the data originates at the local level, but very important types of data come from other levels, including complete information of an issue or topic that transcends jurisdictional boundaries (region or state). Thus there is a growing level of support for policies, interfaces, standards, and relationships that enable government, companies, organisation and citizens to interact and share in the collection and dissemination of geographic information across jurisdictions.

In the Canadian context, public and private-sector partnerships focus on partnering and leveraging the resources of the private sector to accelerate access to spatial data and technology development. GeoConnections, the program responsible for implementing the Canadian Geospatial Data Infrastructure (CGDI) has placed particular emphasis on partnerships between the federal and provincial and territorial governments and the private sector and academia. Programs focus on working across governments, and with stakeholders and the private sector to advance the amount of information accessible through 'clearinghouse' systems, the development of data frameworks to ease data integration, fostering advanced technology and application development, and building supportive policies to speed industry growth. To this end, guiding principles for the provincial and territorial government agencies involved in geomatics have been agreed upon (see box).

Canadian Geospatial Data Infrastructure:
Principles for Data Partnership
(http://www.geoconnections.org/english/partnerships/index.html)

1. Data should be collected once, closest to the source and in the most efficient way possible, with a view towards increasing the vertical integration of the data.
2. Geo-info data should be as seamless as possible, with co-ordination across jurisdictions and boundaries when possible.
3. Data should be collected, processed and maintained according to international standards to maintain data integrity across databases, and to enable the addition of value, further enhancement and easy access and use.
4. Upon agreement, partners should contribute equitably to the costs of collecting and managing the data, and should be allowed to integrate the resulting information into their own databases, for their own use and for further distribution to their stakeholders.
5. There should be an attempt to harmonise terms and conditions for use where practical. In the absence of such agreement, each agency should be free to establish its own terms and conditions for such information.
6. Agreements between agencies will normally be negotiated on a case-by-case bilateral or multi-lateral basis, according to these principles of partnership.
7. Partnerships between agencies should be simple and support the principles of the CGDI, open to the participation of interested stakeholders within any level of government, the education communities or the private sector.
8. A group or agency within each province and within the federal government should be designated to promote and co-ordinate the development of a common

geospatial data infrastructure, both within its jurisdiction and between jurisdictions.
9. CGDI is national in scope, and must meet the needs of a wide range of geospatial user communities, data producers and different areas of the private sector.
10. CGDI must consist of a set of co-ordinated and interrelated policies, practices and possibilities that build on the vision.

Develop clearinghouses and use common standards for data and technology: The technical underpinning of a SDI is a common framework of standards, tools and services based on these standards. In this three-tier model, applications work with metadata and data content and services that exist on the enabling infrastructure. The following technical elements are important components of a SDI:

- quality metadata,
- residence of metadata in on-line directories,
- good data management,
- access to services on-line,
- their documentation in directories and
- reference implementations of software to demonstrate capabilities.

For existing and emerging standards and free- or low-cost software solutions based on these standards, please consult Chapters 2-7. (include links?)

The development of the Portuguese Spatial Data Infrastructure serves as an example of the importance of outreach activities that parallel the implementation of the technical elements of a SDI (see Example 18). The Portuguese SDI differs from other SDI's by having a centralised metadata catalogueue. Usually metadata is organised in a distributed way. Nevertheless, the example demonstrates that in order to gain support for the system (i.e. increasing the number of users of the system), new interfaces were developed according to feedback from the users themselves and through the development of tools that are more devoted to the needs of citizens. The Portuguese experience also shows that a SDI can be developed incrementally with improvements implemented step by step.

Example 18
User involvement in the technical implementation: In 1990 the Portuguese government created SNIG, the Portuguese GI infrastructure, as a national public service (http://snig.cnig.pt). The main goal of SNIG was to ensure the connection of Portuguese users and producers of digital geographic information through a network. This goal implied the development of catalogueues describing the available geographic information. By that time, the majority of public agencies were more concerned with the production and organisation of digital geographic information than with the dissemination process. It was felt that data producers were not prepared to manage their own metadata registers. Thus, the creation and maintenance of the metadata that currently supported the SNIG was organised in a centralised way by its co-ordinator, the National Center for Geographic Information (CNIG). By the end of

1994, taking advantage of the multiple opportunities for publishing data offered by the World Wide Web, the Portuguese GI metadata catalogueues were implemented in a Relational Database Management System and CNIG started to build an HTML interface to allow the query of the metadata and the retrieval of the available data sets. The SNIG network was finally launched on the Internet on May 3, 1995. The main concern was to connect users to the available digital data sets, creating an operational system that could be improved in the following years. Therefore, the metadata catalogueues were not based on any metadata standards. Common sense, some guidelines provided by the CORINE Catalogueue of Data Sources Project and the identification of the main geographic information sources were used to design the database. During this stage, the system structure and design was mainly oriented towards the professional user.

Subsequently, the creation of new metadata catalogueues obliged to rebuild a new WWW interface. While the first SNIG interface was developed without the implementation of formal usability studies, usability testing was required to support further developments of the SNIG. In order to rebuild the SNIG site, qualitative research involving users was carried out for the first time. The research was designed to answer to the following questions:

Which would be SNIG potential users groups?
What geographic information options would users need?
What would the users be looking for in an infrastructure such as SNIG?

The main results of this research pointed out that it would be necessary to develop a friendlier interface. The new interface should adopt informal and non-technical terminology and include search engines by terms and geographic location. The need of more geographic information for non-professional use and the adoption of more common data formats were also stated. It would also be important to include raster images to illustrate the available information. In July 1999, an alternative SNIG user interface was launched (GEOCID). GEOCID is more appealing and information oriented, avoiding complex tasks and navigating routes to access to the data. In addition, new applications were developed based on the information citizens are interested in. An application that allows the user to navigate through Continental Portugal, select specific locations and download the part of the orthophoto he is visualising on screen was developed. The launch of GEOCID was a big success (http://ortos.cnig.pt/ortofotos/ingles/).

Recommendations: Outreach and Capacity Building Options for Implementing a SDI

By overcoming inefficiencies, a coherent and consistent SDI can ensure that geographic information may be used to address complex social, environmental, and economic issues. The following guidelines indicate some of the outreach and capacity building activities that can be used to foster the implementation of a SDI:

- A practical step to take in the development of a national SDI is the development of a vision, detailing a vision of the desired future and a clear sense of how SDI components could serve that future and help to realise it. This also involves setting clear priorities and defining a strategy or policy to accomplish the vision.

A workshop organised with the stakeholders to define and create a national co-ordinating body, considering its structure in terms of an existing or newly created institution, working groups and/or committees. In countries where GIS implementations are highly dependent on donor involvement in terms of funding and technical expertise, donor representatives should be considered as stakeholders and included in the process of building a SDI. The co-ordinating body needs to be mandated to manage the required activities and devise an action plan to co-ordinate the activities.

Consideration needs to be given to the necessary resources for implementing strategy, policy or plans and activities, considering staff, technical know-how, material, and funding opportunities such as innovative partnerships.

Formal working groups should be organised around well defined objectives, strategies, plans, programmes, and actions, and not simply for informal and limited consultations. These working groups would be made up of interested parties and experts to deal with specific aspects of SDI such standards (metadata, exchange), national data sets, policy, clearinghouse and how to assimilate existing technological solutions into the local context.

- Awareness creation of SDI components should be considered down to the lowest level and with strong management support and leadership.
- Plans should be developed and implemented for the dissemination of information on SDI activities under way, including the information about the SDI components, available technical best practices, and the promotion of the use of existing technologies and standards to support the development of a SDI, for example by establishing WWW pages on the Internet or using printed media or CD-ROM where Internet connections are limited.

Measures should be taken to monitor, analyse and participate in, developments at international levels that affect the use of standards and supporting technologies in the national context. This entails assigning clear administrative responsibility for tracking key developments are the international level and within the GSDI community.

Within SDI development, the role of donors should be clarified to support activities by way of following local priorities such as interoperability of different GIS implementations rather than wishing to be associated with a particular type of activity irrespective of cost effectiveness or fit with the broader institutional or national objectives.

References and Linkages

GSDI (1998) "Conference Resolutions, Recommendations and Findings", 3rd Global Spatial Data Infrastructure (GSDI) Conference, Canberra, A.C.T., Australia, 17-19 November.

Gouveia, C., Abreu, J., Neves, N., Henriques, R. G. (1997) "The Portuguese National Infrastructure for Geographical Information: General Description and Challenges for the Future", GISDATA Conference Proceedings.

Henriques R. G., Fonseca, A. *et al.* (1999) "National System for Geographic Information (SNIG): The Portuguese National Infrastructure for Geographic Information", Madame Project: 1st Progress Report.

Mapping Science Committee, Board on Earth Sciences and Resources, Commission on Geosciences Environment and Resources, and National Research Council (1994) "Promoting the National Data Infrastructure Through Partnerships", Washington, D.C.: National Academy Press.

Mapping Science Committee (1993) *Toward a Coordinated Spatial Data Infrastructure for the Nation*, Washington, DC: National Academy Press. Execrative Summary available at http://38.217.229.6/NAPA/NAPAPubs.nsf/00a36275d19681118525651d00620a03/229b79ae768d77e48525658c0061a3bd?OpenDocument).

Masser, I. (1999) "All Shapes and Sizes: The First Generation of National Spatial Data Infrastructures", *International Journal of Geographical Information Science,* Vol. 13 (1), pp. 67-84.

Mbudzi, M., Jairosi, Y., Vogel, D. and Bohnet, D. (1997) "Best Practices on Environmental Information Systems (EIS): The Case of Zimbabwe", Program on Environmental Information Systems in Sub-Saharan Africa, May.

Mendes, M. T., Joaquim, S.P., Hengue, P. and Gerbe, P. (1998) "Best Practices on Environmental Information Systems (EIS): The Case of Mozambique", Program on Environmental Information Systems in Sub-Saharan Africa, May.

Nicolau, R. (1998) "Adoption of the Metadata Standards within SNIG", workshop on "Challenges and Future Developments of GI Infrastructures: The Portuguese Experience", GIS PlaNET'98 Conference, Lisbon, FIL, 7-11 September.

Ryerson, R.A. and Batterham, R.J. (2000) 'An Approach to the Development of a Sustainable National Geomatics Infrastructure', *Photogrammatic Engineering and Remote Sensing,* January, pp 17-28.

United States National Academy of Public Administration (1998) "Geographic Information for the 21st Century: Building a Strategy for the Nation", Executive Summary, January, http://www.napawash.org

Wehn de Montalvo (1999) "Survey of Spatial Data Sharing Perspectives in South Africa - Views on the Exchange of Spatial Data Across Organisational Boundaries", Summary Report, SPRU - Science and Technology Research, University of Sussex, December.

Chapter Nine: Case Studies

Editor: Mark Reichardt, OpenGIS Consortium, Contributors: International: Digital Earth; Regional: Africa (Zimbabwe Harare Team and EIS); and National: Santiago Borrero, Colombia

Introduction

While Chapter 8 has outlined the elements of outreach and capacity building needed to form a viable national and global SDI, this chapter provides some examples of SDI implementations from a national, regional, and global perspective. The documentation of case studies is an effective mechanism to help convey the underlying factors that led to the growth of spatial data infrastructures. This chapter will highlight some of the success stories, shortfalls, and issues that characterise the state of the National and Global Spatial Data Infrastructures today.

Contributors from both developed and developing countries have provided case studies for this chapter. Wherever possible, authors have attempted to cite the major factors leading to success or shortfalls in a particular case study. The reader should note that this chapter will grow to include greater comparative information as more case studies are examined and incorporated. For this first publication of the SDI Cookbook, single national and regional case studies are examined.

Local Case Study – Within nations, localities are increasingly addressing decision-making through the use of geographic information and tools. The ability for spatial data infrastructure to deal with local as well as broader national issues is essential. A case study from the United States involving crime management is highlighted as one of many examples of local communities that are benefiting from the investment in SDI towards improved community service. Our thanks go to Mr. John DeVoe of the US Department of Justice (mailto:john.devoe@usdoj2.gov) and the staff of the Baltimore Police Department for their contributions.

National Case Study - the Colombian experience in developing and harmonizing geographic information systems is examined. Its main purpose is to contribute to identify best practices in Spatial Data Infrastructures as a means to increase geographic information availability, access and use to support making decisions and to promote sustainable development. A team of authors from Colombia's IGAC provide a comprehensive assessment of the Colombian experience in establishing a national SDI. Acknowledgements go to Santiago Borrero Mutis (sborrero.igac.gov.co), Iván Alberto Lizarazo Salcedo (mailto:ilizaraz@igac.gov.co), Dora Inés Rey Martínez (mailto:direy@igac.gov.co), and Martha Ivette Chaparro (mailto:mchaparr@igac.gov.co) for their contributions to this chapter.

Regional Case Study - A case study from the SADC Regional Remote Sensing Unit, which is part of the SADC Regional Food Security Programme, facilitates training programmes and technical support in the field of remote sensing and GIS in support of early warning for food security and natural resources management. This case study is provided as an example of how a focus on critical regional issues yields elements of infrastructure valuable for cooperating nations. Camille A.J. van der Harten (mailto:cvanderharten@fanr-sadc.co.zw), Senior Adviser, SADC Regional Remote Sensing Unit, Harare – Zimbabwe provides an outstanding overview of that effort, its successes, and issues.

Global Case Study - For the global case study, the authors reviewed the major organisations, systems, and processes that are operating to achieve one or more aspects of the Global spatial Data Infrastructure. Although a true GSDI is not a reality today, a review of the current areas of emphasis is in order. Thanks to the members of the Digital Earth Team Tim Foresman (mailto:foresman@umbc.edu) and Gerald Barton, and the University of California Santa Barbara / Global Map team (Jack Estes (mailto:estes@geog.ucsb.edu), Karen Kline (mailto:(kline@geog.ucsb.edu) for their contributions.

Local Case Study

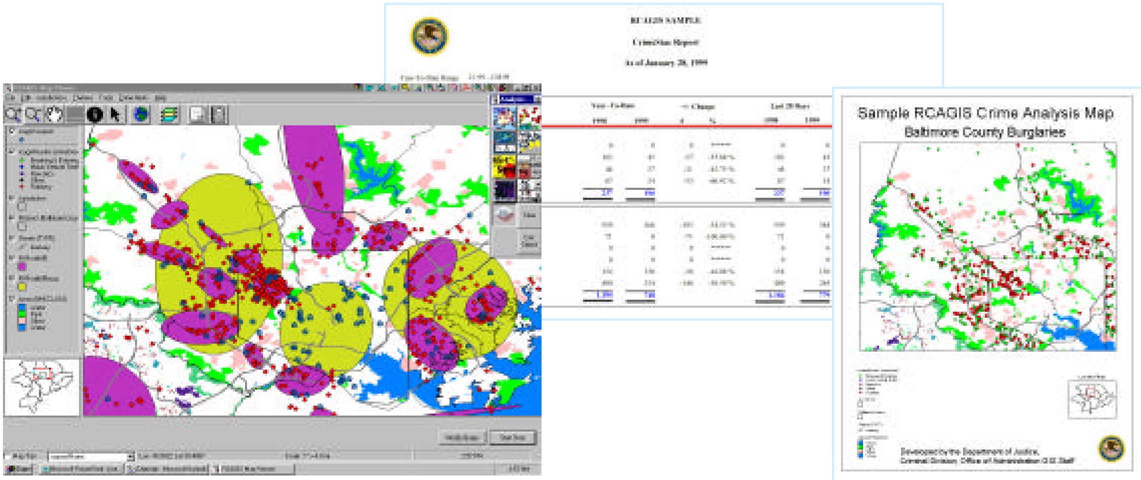
In this chapter, the authors have highlighted some national, regional, and global case studies that are helping to contribute to the GSDI. This section is intended to illustrate one example of the successes at the local level to advance the ability of communities to improve their decision-making through the use of spatial data infrastructure.

Background, Context and Rationale

The reduction of crime in communities across the United States is a major goal to assure safe and liveable communities. Although crime types and rates vary from locality to locality, the use of geographic data and tools is rapidly becoming a key resource to better understand and more effectively deal with crime. In the United States, community safety and policing are primarily functions of local and state governments. Recently, Baltimore City, Baltimore County, and other neighbouring law enforcement organisations realised that cooperative analysis of crime trends regionally would reveal a more complete picture of crime trends. As a result, the City of Baltimore, Baltimore County and other police departments in the Mid-Atlantic area of the United States came together to target and reduce the amount of crime by identifying and implementing methods to standardise their approach to the management and use of crime data and related geospatial information.

Organisational Approach

In the early 1990's, the United States Department of Justice, recognizing the value of geospatial data and techniques in managing crime, established partnerships with local law enforcement organisations to illustrate the value of GIS applications in the identification, visualisation and analysis of crime trends locally and regionally. These early partnerships were also designed to show industry the potential market for applications to better address crime management. The success of these early efforts, led to the creation of larger regional partnerships to address crime issues using geospatial data and geospatial applications. Law enforcement organisations working collaboratively in the region helped the US Department of Justice develop the requirements for a Regional Crime Analysis GIS application. Participating communities agreed to use the RCAGIS crime mapping, analysis, and reporting applications developed under contract by the Criminal Division, United States Department of Justice. Additionally, Baltimore City applied for and received designation of this effort as a National Spatial Data Infrastructure (NSDI) Community Demonstration Project. This designation brought in additional support from the Federal Geographic Data Committee and the Vice President's National Partnership for Reinventing Government.



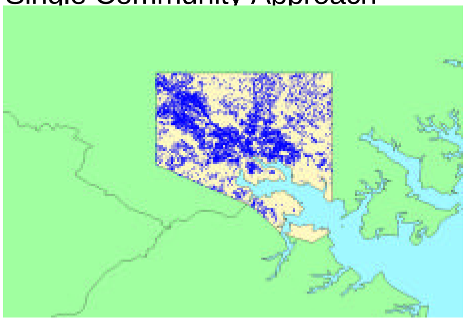
Implementation Approach

The Regional Crime Analysis GIS (RCAGIS) was developed to provide police officers, crime analysts, investigators, chiefs/commissioners/sheriffs, and managers a powerful, yet easy to use crime mapping, analysis and reporting application. RCAGIS is designed to assist police departments in their tactical and strategic responses to crime and to help create an environment where police department personnel assume responsibility for increases and decreases in the amount of crime. RCAGIS operates in a PC environment and uses ESRI's MapObjects. MapObjects was chosen because it is relatively inexpensive to implement on a moderate to wide-spread basis..

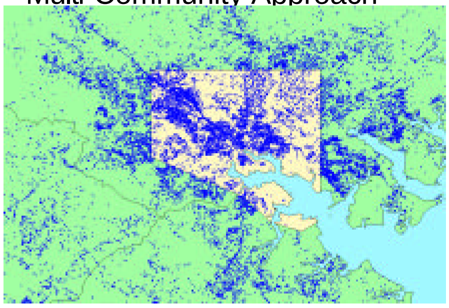
RCAGIS seamlessly integrates CrimeStat, a very powerful spatial statistic tool developed by Dr. Ned Levine, of Ned Levine and Associates. The RCAGIS programming code is available, free of charge, through the Criminal Division, United States Department of Justice homepage (http://www.usdoj.gov/criminal/gis). . Through this cooperative partnership, the DOJ and local police departments in the

Baltimore – Washington area standardised the format for crime incident data and the methods of mapping, reporting, and analysing crime.

Single Community Approach



Multi-Community Approach



With the success of the RCAGIS program comes the need to address how to manage the growing volume of geographic data that are produced by police departments or other local government agencies in the region. Through support from the US Federal Geographic Data Committee, and the designation of Baltimore as a NSDI Community Demonstration Project, training and technical assistance was provided to the Baltimore City Police Department to implement metadata standards and practices. Additionally, spatial data clearinghouse nodes will be established to inventory and advertise Baltimore City Police Department's designated geographic data. The posting of metadata allows the law enforcement community to know what geographic data is available in the area. Additionally, metadata and clearinghouses can accommodate both public access to data, and the management of data restricted only for law enforcement use due to local policy.

The RCAGIS program has helped localities to improve collaboration on issues of mutual importance. The program illustrates to law enforcement staff the value of metadata and clearinghouses in improving the ability to inventory and share information. By standardizing data elements and the metadata that describes this data, law enforcement organisations have improved their ability to communicate issues across jurisdictional boundaries, see the broader implications of crime, and devise more comprehensive solutions to apprehend offenders, and to reduce crime trends overall. Finally, by using clearinghouse resources, law enforcement will be able to discover and apply additional environmental, social, and economic data sets to enhance police departments' crime analysis and tactical and strategic responses to crime, thereby reducing the amount of crime and residents' fear of crime in our communities.

Recommendations

Establish expanded partnerships - A broad multi-jurisdictional view of crime is often necessary to understand overall crime trends. Indeed issues related to crime, the environment and the economy are not typically contained within community boundaries. Partnerships and collaboration through the sharing of data, standards and processes enhances the ability to understand and manage patterns of crime that are significant to the larger area. Partnerships with the federal government provided expertise, training to deal with many issues, and also provided some funding to advance this effort as well.

Educate spatial data managers and users on the value of SDI practices – metadata, clearinghouses, and standardisation are concepts that until recently were very unfamiliar to the law enforcement community, and will not be readily adopted unless the appropriate level of education and outreach is applied to illustrate the value of metadata and standardisation to assure data accessibility, quality, availability, and overall management.

National Case Study – Colombia

Background, Context, and Rationale

As with many nations around the world, the major drivers for geographic information infrastructure in Colombia stem from the nation's programs for governance to address national issues related to the environment, the economy, and social issues. Drivers also include private sector interests in the major areas of Colombia's Economy. Furthermore, Colombia understands that issues of national importance often extend beyond its borders, so the growth of the national infrastructure must accommodate collaboration regionally and potentially globally. This case study will focus on the efforts of Colombia to establish a national SDI, and discusses the steps Colombia has taken to assure SDI compatibility to address regional and global issues such as those raised via the UN Agenda 21.

Initiatives to coordinate SDI actions in Colombia at a national level face significant constraints like decreasing budgets, inter-organisational barriers, lack of high level support, limited capacity for research and development and lack of knowledge about the geographic information market, among others. Despite these restrictions, experience has shown that specific steps to define and implement a national geographic information strategy can be accomplished, providing that government agencies decide to work together, reduce costs, avoid duplication of efforts, and recognise the role that the private sector and academia can play. User demands can trigger the necessary partnerships and alliances to produce and share information.

The Colombian Spatial Data Infrastructure (ICDE) is defined as the set of policies, standards, organisations and technology working together to produce, share, and use geographic information about Colombia in order to support national sustainable development. The ICDE is a young but promising, initiative. The lessons learned through its design and development may be useful. Due to the fact that it lacks a formal mandate to build the Colombian NSDI (as compared to the U.S. case), the ICDE has followed an empirical approach, in which design and development are not completely separated and well-defined stages are utilised. The ICDE has struggled to gain visibility and support while under pressure to show results.

The ICDE must be understood as an initiative that is under construction, in which practice is used to refine the concepts. Various government organisations, private companies, and universities are laying the ICDE building blocks. The IGAC, DANE, IDEAM, INGEOMINAS, ECOPETROL and the Ministry of the Environment, among others, have made valuable contributions. While work on standards and data production has been remarkable, yet still insufficient, reaching agreements on policies and high-level support seems to be the major area requiring further efforts. This document explains why the ICDE, the Colombian NSDI, was born and how its family is taking care of it and helping it to grow.

Quick Overview of Colombia

The republic of Colombia, located in northwestern South America, encompasses a total area of 2,070,408 square kilometres, of which 1,141,748 are on the mainland. In 1992, the population of Colombia was approximately 36.2 million people. The country is a rich mix of peoples, including Mestizo (European-Indian), European, African-European, African, African-Indian, and Indian descent. The main language in Colombia is Spanish, but over 200 indigenous Indian languages are also spoken.

Colombia has a democratic political system and Santa Fe de Bogotá is the capital city. The major industries are textile production, coffee, oil, sugar cane and food processing. The GDP in Colombia is US $172 billion. Inflation currently runs at about ten percent.

Colombia is the fourth-largest country in South America and the only one with coasts on both the Pacific and Caribbean oceans. It shares borders with Panama (to the northwest), Venezuela (east), Brazil (southeast), Peru (south) and Ecuador (southwest). The Colombian territory also includes the San Andrés and Providencia island groups, 700 km. (435 m.) northwest of the mainland, in the Caribbean Sea. The archipelagoes are 230 km. (140 m.) east of Nicaragua.

Three Andean mountain ranges run north and south through the western half of the country (about 45% of the total territory.) The eastern sector is a vast lowland, which can be generally divided into two regions: a huge, open savannah in the north, and the Amazon in the south (approximately 400,000 sq. km.).

Colombia has the highest number of plant and animal species per unit area of any country in the world. The country's network of reserves includes 33 national parks, six small reserves, known as *"santuarios de flora y fauna"*, two national reserves and one special natural area. Their combined area constitutes 7.9% of Colombian territory.

Geographic Information in Colombia

Most geographic information on the Colombian territory is produced by government agencies that have specific mandates. The DANE is responsible for conducting the census, both social and economic. The IDEAM is in charge of hydrology, meteorology, and environmental studies. The INGEOMINAS works in the area of geoscience, environmental mining, and nuclear energy. The IGAC carries out topographic mapping, cadastre, soil, and geographic activities. All these institutes are very experienced in their respective areas, both in terms of the time they have put in, as well as the amount of valuable information produced throughout the country. Over the past decade, pursuant to presidential decrees, these Colombian agencies have developed modernisation processes for structural and resource reorganisation in order to fulfill their institutional goals and the community's needs. New technology has been incorporated into the production flow, people have been trained, and agencies are furnishing digital products to users.

Aside from the above-mentioned agencies, some companies share a small, but increasing, portion of the geographic information market. They provide products and services to the government and private sector, and help to produce topographic and thematic mapping and develop GIS applications.

In the 1990's, an awareness of the benefits of geographic information started to grow among municipalities, environmental agencies, oil companies, and the utilities sector. Seeking to meet legal requirements[6] or business challenges, some people turned their eyes toward geographic data. A demand for digital base maps was born and grew quickly, although not always supported by adequate funding. It has taken time to convince users that the government cannot provide new digital products for the low cost of duplication, as for analogue information.

Unfortunately, high-level government decisions currently do not benefit from geographic information. Despite increasing recognition of its role to generate knowledge, provide added value to identify problems, assist in proposing alternatives and defining a course of action, geographic information discovery, access and use have not spread as desired. Indeed, government agencies face budgeting constraints for the funding of production and maintenance of their databases. In most cases, government agencies must attempt to find ways to accomplish their principal functions and achieve a minimum level of cost recovery.

[6] According to recent legislation (Ley 388 de 1997), municipalities must set out a territorial ordering plan to define and regulate land use. Geographic data are key to ensure compliance with the law.

National GIS Projects

With a view to fulfilling their mandates, government agencies are carrying out various initiatives to develop national information systems in the areas under their jurisdiction.

Environment Information System for Colombia (SIAC) - According to *Law 99 of 1993* and *Decrees 1277, 1600 and 1603 of 1994,* the Ministry of the Environment shall lead the coordination of the National Environment Information System (SINA) and establish the Environment Information System (SIA), and the IDEAM shall manage the implementation and operation of the SIA and advise the CARs[7] to do the same in their areas. Other research institutions (INVEMAR, SINCHI, John Von Neumann, Alexander Von Humboldt) shall contribute to system implementation throughout the national territory with the aim of providing timely and sufficient environmental information to support policies and decision-making.

At the provincial level, some CARs have also developed environment information systems, most of them successfully. However, these various developments lack convergence and coordination.

At the present time, the Ministry of the Environment is initiating a system for the planning, design, and development process, to harmonise efforts and strengthen and consolidate the SIAC. This system targets water resources, pursuant to the National Environmental Policy, which establishes water as its principal focus. The Policy also involves the community in the development strategy through their participation in the area of information appropriation.

The National Environment Information System (SINA) – The IDEAM has developed the SINA's basic module and provides information in real-time about environmental status and changes. Some of its products are: *Environment in Colombia, Natural and Socio-economic Impacts due to the Pacific Hot and Cold Phenomena – el Niño and la Niña, The National Water Study, Offer-Demand Relations and Sustainability Conditions, Vegetation and Land Use, Morphogenic Systems and the Stability of the Geological Morphostructure and Superficial Formations.*

The National Geoscientific Information System (SING) - Under *Decree 1129 of 1999,* the INGEOMINAS shall conduct research and generate basic information for geoscientific knowledge and the improvement of the Colombian subsoil. To this end, the INGEOMINAS shall survey, obtain, compile, integrate, validate and provide in digital and standardised format, subsoil information, including geology, geophysics,

---

[7] Regional Autonomous Corporations are environmental administrative units in charge of the management of renewable natural resources and sustainable development in their jurisdiction (major river watersheds).

geochemistry, geomechanics, non-renewable resources and geology-based hazard monitoring. The INGEOMINAS will develop the SING as an integral part of the Colombian Geographic Information System.

The INGEOMINAS has produced several digital atlases over the past few years, in the areas of geology, geochemistry, gravimetry, geological hazards, metallogenesis, geochemical anomalies and mining activity.

The National Geostatistical Information System (SAIG) – According to Decree 2118 of 1992, the DANE shall manage the SAIG. The SAIG fosters the integration of social, demographic and economic statistical information obtained from census taking, surveys and administrative records, using current technology to store, query and analyse information.

The SAIG engages in the following tasks: design and methodology for census-taking, surveys and research on social and economic data, such as quality of life, construction, national home surveys, the consumer price index, national population and housing census, and collection of information for planning; development and control. Other tasks include definition and updating samples, processing information, analysis, and publication of results.

The National Geostatistical Framework links statistical information with the corresponding geographic sites. It is made up by political / administrative groups and geographic sectors oriented toward statistical activities. It attempts to improve social welfare, sustainable development and Colombia's competitiveness.

The IGAC Geographic Information System (SIGAC) - *Decree 2113 of 1992* empowers the IGAC to draft and update the Official Map of the Republic of Colombia, develop policy, and undertake national government programs in cartography, agrology, cadastre and geography. This is done through the production, analysis and distribution of geo-referenced environmental and cadastral information, which is aimed at supporting planning and territorial ordering processes.

The IGAC has developed the Integrated Geographic Information System, which is designed to build and maintain national digital databases in topography, soils and cadastre. It began to be implemented in 1995.

The conceptual model of the IGAC's (SIGAC) Integrated Geographic Information System included the following aspects:

- Design and implementation of an integrated Data Model for the 1:2,000 and 1:25,000 scales. In this model, the real world is represented by a Digital Landscape Model (DLM) (primary model), where the different objects are classified, coded and transformed through a cartographic work into a secondary model, the Digital Cartographic Model. The objects are categorised in terms of themes, groups, and object classes.

- Creation of the Spatial Database according to the Data Model. The data structure simplifies spatial analysis and linkage of geographic objects to external data in order to be available for multipurpose use. The topographic data are entered into the system using the analytical restitution of the photos. Digitising the existing maps captures the cadastral and soil information. The SIGAC structure and content includes: land fixed points, photogrammetric fixed points, land transport, aerial transport, shipping transport, engineering structures, vegetation, water streams, water bodies, relief, buildings, land ownership and territorial and administrative boundaries.
- Establishment of data–exchange formats for internal and external users of the system.
- Definition and establishment of standards.

Some of the main tasks performed by the SIGAC are: calculations, surface intersections, interpolations and topographic modelling; land registration; land valuation; production of soil homogeneous zones; derivation of physical and geo-economic homogeneous zones; and production of land use maps. The principal products supplied by the SIGAC are: topographic maps at different scales, cadastral maps, soil maps, land registration certificates, land use maps, physical homogeneous zone maps, geo-economic homogeneous zone maps, land homogeneous areas for cadastral purposes, land suitability classification maps, land capability classification maps, digital terrain models, and statistical information regarding buildings, parcels, owners, etc.

Until now, the IGAC has made great efforts to bridge the gap in basic map availability and currency. Coping with adverse meteorological conditions and taking advantage of the new geo-information technologies, the IGAC is trying out new data sources, procedures, and products. Despite some achievements, more R&D is still needed. A great deal of topographic and cadastral digital maps have been produced, focused on 1:2.000 scales for cities and towns and 1:100.000 scales for rural areas.

The National Oil Company Information Infrastructure (GEODATA) - Recognizing that the current manner of conducting the oil business in Colombia is too expensive and time consuming, ECOPETROL has entrusted the ICP (its research center) with the task of defining policies and standards and developing an infrastructure to manage geographic information, according to new technologies and customised to company needs. Its most ambitious project has been the development of a distributed data repository to provide a common, high-quality warehouse for Colombian primary and interpreted petrotechnical data. The data warehouse will ultimately aim to be Colombia's official repository for petrotechnical data on oil exploration and production. Primary petrotechnical data includes all non-interpretative data that may be used by the industry in its day-to-day work.

The Coffee Information System (SICA): The Colombian Coffee Growers Federation (FEDERACAFE) is a non-profit institution. It was created in June of 1927 and currently unites almost 300,000 producers.

The FEDERACAFE has developed strategic plans to improve the competitiveness of Colombian coffee and to provide research and development programs on improved technologies for production, the post-harvest process, coffee quality, the management capacity of coffee producers, and marketing to increase the demand for Colombian coffee.

One of the programs that has been developed is the Coffee Information System (SICA). This system permits the coffee authorities, the Federation and the producers to base their work on strategic and updated information that allows them to design policies and programs to improve competitiveness, the sustainable development of Colombian coffee production, and the welfare of the coffee producers.

The SICA includes the following elements:

The coffee plantation structure (plots, areas, number of plants, varieties, borders, brightness, meters above sea level).
Socioeconomic aspects of coffee growers and their housing.

The Federation has developed a specialised Software "SICA" or AFIC (Attention for Farms and Coffee Growers).

Despite the developments described above, it is clear that each institution has built its information systems independently and that national policies and guidelines were non-existent at the time they started these processes. Due to this, interorganisational links have not been strengthened as needed, the roles of the agencies have not been clarified, and analogue-digital data conversion activities may have been duplicated. Digital databases were built autonomously and problems soon arose: data were out-of-date and incomplete, heterogeneous in content and quality, poorly documented, hard to find and difficult to integrate. Client needs were not recognised as required. An awareness of these problems led to the need for standardisation.

First steps towards a national geographic information strategy

The IGAC, which is in charge of the national databases on topography, cadastre, soil and geography, developed in 1995 a geographic object classification scheme for use in different scales. Other institutions adopted the IGAC scheme and added their own objects. This was the first step to achieving order in-house. Around the same time, ECOPETROL, the national oil company, started its project Geodata, which focused on geographic data standards and metadata. Both initiatives pushed forward the creation of a national committee in charge of defining geographic information standards. Under the auspices of ICONTEC, the Colombian body for standardisation and certification, and with coordination by the IGAC, more than thirty entities from government, the private sector and academia contribute to this committee. Until now, efforts have been concentrated on geographic metadata, basic object catalogueuing, quality, and terminology.

As user understanding of GIS capabilities grew, an understanding of the need for homogeneous and consistent data also grew. Government agencies began to understand their role was changing: they had to become information providers and not only data producers. Private companies started to share an emerging digital geographic information market. Partnerships developed to produce and update topographic and cadastral data. The IGAC and other institutions convinced some city authorities to fund digital database projects on a fifty-fifty cost sharing basis between municipalities and the Colombian government. The results demonstrated the benefits of sharing costs and information.

However, interorganisational cooperation alone could not accomplish SDI objectives, nor would it be done by Colombian agencies acting alone, without broader participation by industry, academia and local governments. Cooperative efforts would have to be augmented by national policies and guidelines to clarify the roles, responsibilities, priorities, and legal issues, such as copyright, prices, liability and custodianship.

A high level team drafted some government policies on information in 1996[8], producing policies that emphasised the need to manage information like a strategic national resource. These policies viewed the use of information technology as a means to promote social welfare and citizen service, and to link government agencies with outside sectors. Nonetheless, specific policies on geographic information were still missing.

As a consequence of the above, geographic information availability and access were not optimal. Furthermore, geographic information was not being used to its full potential for decision-making and to support sustainable development. A national strategy for geographic information was needed, to focus on the following priorities:

Definition of basic policies.
Production of fundamental data.
Documentation of geographic data.
Improving access to users.
Education and consciousness raising.

Subsequently, the ICDE concept was born in late 1995. The ICDE was influenced by American and European concepts yet retained a local flavour. This local flavour was required to address unique Colombian characteristics: a developing country and government, a nation rich in biodiversity, mineral resources, natural hazards and socioeconomic problems, and the Andean region, which is challenging to map due to meteorological conditions. Early success in the standardisation work done by

[8] Políticas de tecnología informática para el sector público colombiano *("IT Policies for the Colombian Public Sector")*, DNP, COLCIENCIAS, DANE, 1996.

technical teams and increasing demands by government users to account for programs using national information encouraged public agencies to deal with the remaining issues.

Organisational Approach

In 1998, the Colombian government defined as a priority the establishment of a long-term multilateral alliance between Colombia and The United States, the "Environmental Alliance for Colombia" *(Alianza Ambiental por Colombia)*, aimed at the promotion of technical, scientific, managerial, informational, financial and political cooperation for the knowledge, conservation and sustainable development of Colombian natural resources[9]. The Alliance's mission and priorities include:

Management of ecosystems
Cleaner production
Environment Information System
Supply and demand of environment products and services
Water

A round table was set up on each of the above issues under the aegis of the Ministry of the Environment. The Directors of the IGAC, DANE, INGEOMINAS and IDEAM were called upon to participate and coordinate actions to support decisions on the environment. The discussion quickly moved to the need to strengthen interorganisational links, increase information production and sharing, improve the status given by the Colombian government to geographic information, and define a national geographic information strategy.

In November 1998, an Inter-Institutional Committee was set up to create consensus on different topics. The government agencies in charge of geographic information production agreed to work jointly to define policies, guidelines and strategies to foster the production and publication of geographic data in Colombia and facilitate data integration, use and analysis by the agencies' information systems[10]. The committee also decided to promote carrying out actions to develop autonomous information systems in a coordinated and harmonised way as integral part of a national geographic information system. The Committee agreed to coordinate actions in the following areas:

Definition of guidelines and strategies to produce, process and make available geographic information.
Definition of products under the aegis of each agency, taking into account user needs.

---

[9] In October, 1998, in Washington, Colombian President Andrés Pastrana officially launched the *Alianza Ambiental por Colombia.*

[10] *Document: "Proposal for the Design and Implementation of a Colombian Geospatial Information System" (Cartagena, May 6 & 7, 1999)*

Strategies for standardisation of products/processes.
Strategies for the development of telecommunications and information technology infrastructure.
Legal and business strategies.
Organisational strategy and roles to develop the Colombian Geographic Information System (ICDE).
Strategies to build the National Geographic Information Network.
Communication and marketing.

The Organisational Strategy will define the actions to be carried out by different agencies in order to implement the agreements on internal structure, organisational culture, and technical infrastructure. The Organisational Strategy will define a clear outline of the responsibilities of each agency in the development and implementation of the ICDE including: interaction, mechanisms for the joint development of projects, and linking to other public and private institutions.

As noted above, action by the Ministry of the Environment, and its viewpoint as a user, triggered the first interorganisational meetings and helped diminish some communication barriers. Major government producers continued to look for better ways to interact and gained valuable insights. However, their collective desire to produce a document with organisational strategies by the end of 1999 could not be achieved. The restructuring process of state institutions that the Colombian government began in mid-1999 focused the agencies' attention inwards, as they struggled against functional instability and turned the inter-agency activity to their own operational issues[11].

Some government agencies that are major users of geographic information, like ECOPETROL, FEDERACAFE and EEPPM, are very interested in playing a role in ICDE development. Indeed, contributions by them to standardisation and their investment in production and updating basic geographic data projects have been valuable. Some have suggested that they attend the next meeting of the Inter-Institutional Committee to enrich the process and widen the scope of the initiative.

In addition, "spontaneous" regional interorganisational initiatives are emerging. Two noteworthy cases are the Aburra Valley Geographic Information System (SIGMA) and Bucaramanga Tecnópolis – Ciudad Digital (the geographic information system for the Metropolitan Area of Bucaramanga). In both cases, municipal authorities and utility companies (water, sewage system, natural gas, telephone, power) agreed to jointly plan, gather and update basic geographic information to support local decision-making. Major geographic data producers have been invited to support technical definitions, but they are not the project leaders.

[11] In the first quarter of 1999, the Colombian President was authorized by Congress to remove, join and restructure state agencies. The deadline was June 1999. Among other reforms, the IGAC was reassigned to the DANE. Nevertheless, the Constitutional Court recently declared these government decisions unconstitutional. Functional uncertainty continues.

Implementation - Approach

Components of the ICDE

The Colombian Spatial Data Infrastructure (ICDE) is defined as the set of policies, standards, organisations, and technology working together to produce, shares, and uses geographic information on Colombia in order to support national sustainable development. Main ICDE components may be defined as: administrative information policies and guidelines, geographic information standards including metadata, fundamental data (framework), and a national geographic information network.

The ICDE has been oriented to addressing development on a priority basis, initially emphasizing two basic areas:
Production and documentation of fundamental data (framework): Linkage of efforts and resources from different institutions, taking advantage of IT, fulfilling standards and user-oriented product technical specifications and focusing on national priorities and programs.
Development of mechanisms to increase access to data and use by the community: Facilitation of metadata queries, data discovery, and recovery. In order to achieve this, development of a legal framework defining both producer and user rights and duties, i.e. copyright, liability, access, and privacy. Two factors are relevant to this effort:
Building the national metadata repository and linking distributed metadatabases via the INTERNET.
Development of the national geographic information network to promote the availability of geographic information products and services.

Implementation of the ICDE

Progress

With respect to the implementation of the ICDE components, the major agreements to date include the following:
Government data producers have agreed to coordinate gathering seamless digital basic databases covering the whole Colombian territory, prioritised as follows:
1:100,000 scale
1:500,000 scale
1:25,000 scale
Some projects are being developed jointly by the IGAC and other institutions using partnerships, which share the costs (through joint investment) and benefits of producing and updating maps, cadastral information, and soil and agrology information.
A national geographic metadata standard was defined in March, 1999 (Norma Técnica Colombiana NTC4611), based on ISO/TC211 and FGDC work. Major

producers have started to document their data sets according to this standard. The ICP, with the assistance of NCGIA-UCSB, has developed metadata and clearinghouse node software tools and has decided to distribute these nationally as a means to stimulate document acquisition, storing and queries. Significant attention is being given to education and training, since it has not been easy to convince people to add a new process (documentation) to the production line. The difficulties encountered in implementing the process have led to the definition of *"minimum metadata"* as alternative to the complete standard.
Other issues are being discussed in the standardisation process:
Quality of geographic information.
Object cataloguеue for basic geographic data.
Satellite geopositioning.
Geosciences.
Terminology.

Government producers have improved their communication and technology infrastructure. For example, Internet WEB sites have been developed for each institution. (For more information, please access their pages: ***ECOPETROL-ICP:*** www.ecopetrol.com.co, ***DANE:*** www.dane.gov.co, ***IGAC:*** www.igac.gov.co, ***INGEOMINAS:*** www.ingeomin.gov.co, ***IDEAM:*** www.ideam.gov.co, ***MINAMBIENTE:*** www.minambiente.gov.co, ***FEDERACAFE:*** www.cafedecolombia.com). Information services are being developed and implemented and GIS online pilot projects are starting. However, keeping in mind that large sectors of the Colombian community have not yet linked to the INTERNET[12], the major agencies continue to develop traditional paper and hard copy products.

Currently, the private sector is involved in helping to produce and/or update geographic data for the Colombian NSDI, or when a government agency decides to hire a firm to publicise some part of the data collection's work. Out-sourced work is estimated to account for about 50% of the total. The commercial sector is also being hired by national and local government to install, operate, and maintain their network infrastructure (cabling, routers, switches, etc.) and/or to disseminate data. Until now, the private sector has not produced or publicised geographic data to a larger public at any charge, but it seems probable that this will occur in the near future.

In terms of the need for international cooperation, the first ICDE project has been defined by the Inter-Institutional Committee and is to be considered by the American Government in the framework of the *Environmental Alliance for Colombia*[13]. The estimated time for the project is three (3) years. It focuses on improving the ability of

[12] 23 people of every 1000 had access to computers in Colombia in 1996 (*Knowledge for Development*, World Bank, 1998-1999).

[13] This project was proposed to the U.S. delegation to the Environmental Alliance for Colombia meeting in Cartagena, on May 6, 1999. An agreement between Colombia and the USA has not yet been achieved.

institutions to effectively support policy formulation and decision-making on environmental issues, within the general framework of supporting sustainability in national development. The project has three components:
Production of national basic cartography (1:100.000 scale).
Development and strengthening of a national geo-spatial information network.
Strengthening institutional skills for the generation of integrated environment information services.
The project's total budget is about US $32,000,000.00. This amount would be funded by national investment and international support.

Issues

Although significant progress has been made, many issues remain to be addressed in order to accelerate the implementation of the ICDE:

Organisational issues: There is no formal mandate to build the ICDE and an institution with the authority to lead the process. Informal initiatives fail to break interorganisational barriers and do not encourage broader participation. Furthermore, institutions continue to focus on the development of geographic information suitable for their own needs and thus, it becomes difficult and costly for other users to "reuse" geographic data.

Policy Issues: There are no formal agreements or processes underway to address privacy, access, use, pricing, and liability. Agencies have autonomous approaches to these subjects, especially in the areas of pricing and copyright. In practice, digital geographic data sets are sold off-line on a single-license basis at prices ranging from 1% to 5% of the production cost. Analogue data sets (photos or maps on paper) are sold at the cost of duplicating them. Private firms mainly produce customised geographic data and charge their clients about 130% of the production cost. In general, this type of data is not available to the public.

User Needs: A user needs study does not exist. A survey of this type would assist in better focussing the efforts and priorities of the ICDE.

Cost-Benefit Study: Similarly, little information is available in Colombia regarding the costs and benefits of geographic data in decision-making. This information is essential to demonstrate clearly the benefit of joining to implement the ICDE to government, business, and citizens.

Conclusions

In developing countries, government agencies in charge of geographic information have the combined challenge of improving performance, learning to cooperate through partnerships within the limitation of budget restrictions, and satisfying increasing user demands. Otherwise, they will be unable to accomplish their goal of providing valuable information to support increased knowledge and national policy. A national spatial data infrastructure initiative seems to be the most suitable strategy to promote long-term multi-sector alliances, not only among government agencies, but also with the private sector and academia, so that all the stakeholders win.

The Colombian Spatial Data Infrastructure (ICDE) is a sound initiative for the promotion of geographic information production with national coverage that will encourage mass use by society and improve sustainable development. Some achievements have been attained and interorganisational barriers are being broken. The ICDE "empirical" approach has been the way to cope with a challenging context and to gain consensus while demonstrating the practical benefits. Nonetheless, the time has come to gain high-level support. The incipient partnerships must be strengthened and coordinated. It appears clear that it is necessary to establish a national geographic information coordination center with a national mandate to guarantee that all participants continue in the right direction.

Positive results should encourage the ICDE stakeholders to renew their efforts, taking into account that initial success depends on the following:

Management: Major producers and users of geographic information must be in charge of running the initiative in a coordinated way and based on national needs. A framework for information management must be established as a key principle to govern interorganisational behaviour.
Participation: A very large number of public and private institutions, non-governmental organisations, academic groups and research centers, or think tanks, must be included. A careful and user-oriented cost-benefit study must be undertaken.
Support: The ICDE must find support from government at high levels to ensure the necessary definitions and funds for the project.
Technical cooperation: The ICDE must be based on lessons learned from most advanced NSDIs, and should be linked strongly to regional and global initiatives to ensure that nations can jointly address issues extending beyond national boundaries.
Research and Development: Appropriate technology needs to be adopted or adjusted through research and development activities.

Recommendations

Seek and acquire high-level government support for the national SDI. The ICDE development process must be accompanied by high-level government support, such as a presidential decree or Ministerial Council Order. Otherwise, the momentum of the Colombian agencies alone will not be sufficient to keep the engines moving for very long.

Define national guidelines for managing geographic information, not only for use in government, but also where this involves the private sector and academia.

When defining basic agreements to stimulate cooperation and focus efforts for the National SDI, these topics must be addressed:

Agreement on the definition of the National SDI.
Clarification of the objectives.
Agreement on the key principles, rules and responsibilities.
Coordinating body
Role of each organisation
Basic policies and guidelines for managing and sharing information
Funding

Early on, develop the first stage national geographic information network through the use of internationally compatible standards and practices. Given that the ICDE is a long term, ambitious project, efforts must be concentrated on developing the first phase of the Colombian geographic information network: a metadata-based clearinghouse, in order to achieve the National Directory of Geographic Information. With a national geographic metadata standard defined and with the development and testing of some customised metadata software tools, the Colombian producers of geographic information now have the challenge of making decisions on documenting their data sets and setting clearinghouse nodes. "Actions speak louder than words".

Study for the Southern African Development Community (SADC) region

Background, Context, and Rationale

A compatible SDI can encourage region-wide collaboration on issues that often disregard national boundaries. While formal regional SDI initiatives are just recently in the discussion or formation stages, there are a number of illustrations of how a regional Spatial Data Infrastructure approach can make a positive difference in dealing with often-difficult issues such as food security. The Permanent Committee on Geographic Infrastructure for Asia & the Pacific is just one example of a regional SDI implementation addressing the joint spatial issues of member nations.

The Southern African Development Community (SADC), which was established in 1980 as SADCC, is promoting regional cooperation in economic development. SADC member nations include: Angola, Botswana, Democratic Republic of Congo, Lesotho, Malawi, Mozambique, Mauritius, Namibia, Seychelles, South Africa, Swaziland, Tanzania, Zambia, and Zimbabwe. SADC has adopted a Programme of Action covering cooperation in various sectors, including food, agriculture and natural resources management. Its secretariat is formed by the Food, Agriculture and Natural Resources (FANR) Development Unit in Harare, Zimbabwe. To effectively address the issues of early warning for food security and natural resource management, a regional spatial database has been developed to assure the timely collection, management and dissemination of critical information and knowledge to the SADC Member-States and other stakeholders.

The SADC Regional Remote Sensing Unit (RRSU) started as the Regional Remote Sensing Project (RRSP) in 1988 and received technical assistance from the Food and Agriculture Organisation (FAO) of the United Nations and financial support from the Governments of Japan and the Netherlands. The technical assistance from the FAO came to an end in June 1998 and since that time SADC RRSU has been gradually integrated in the organisational structure of the SADC FANR Development Unit. The RRSU is financed by the SADC Member States and receives additional financial and technical assistance through a bilateral agreement between SADC and the Government of the Netherlands. The RRSU is a centre of technical expertise, which can facilitate training programmes and technical support in the field of remote sensing and GIS in support of early warning for food security and natural resources management. On an operational basis the RRSU is using low resolution high temporal satellite information to produce information products on rainfall occurrence and vegetation development which is being distributed through the Regional and National Early Warning Units, but also through its own publications, reports and web-site. A variety of training courses and national and regional workshops are organised to create a core of trained experts in the SADC region. An important activity of the RRSU is the development of spatial databases, which are being distributed on CD. The RRSU database includes at present all the basic thematic information (administrative boundaries, infrastructure, land cover hydrology, soils, elevation etc.), as well as the satellite image archive, agricultural statistics, and climate information. In order to develop these information systems further, the RRSU has strategic partnerships with a number of institutes in the SADC region, but also in Europe and the USA. The RRSU spatial database is recognised as a regional (and often a national) standard, and because of this the RRSU is a recognised partner in a number of EIS related activities in the SADC region. At a regional level the RRSU collaborates with the South African National Spatial Information Framework (NSIF) on the development of metadata, which will have a regional outlook.

Genesis of the Regional Early Warning Infrastructure

From the time of its establishment, the RRSU has been working on the use of satellite information to monitor rainfall occurrence and vegetation development in support of early warning for food security. The satellite data covers the whole SADC region and the operational pixel size of the raster images is 7.6 km. With the increased use of GIS technology and the availability of ever-faster computers and more user-friendly GIS software programs, there was a need to harmonise and standardise spatial data sets, not only the raster satellite images, but also the vector database.

In the early nineties most digital spatial data available in the SADC countries originated from small projects. Spatial data available from the Surveyor General Offices was often not in digital format, or in an inaccessible digital format. As a result, many Government offices, small projects, universities, NGO's, started to digitise their own spatial databases.

One of the tasks of the RRSU is to introduce GIS technology. The main problem it faced during the introduction of this technology in the region was the lack of a consistent spatial database for the SADC region. For example, national and sub-national administrative boundaries hardly existed in digital format, or were incomplete. For existing data, there was no cross-boundary compatibility. Other data on infrastructure, basic land use, and hydrology did not exist or was scarce. A soil map had been prepared for a number of SADC countries, but the digital format used made it impossible to use the data for further GIS analyses. The satellite images in raster format from the Meteosat satellite (for climate monitoring) and NOAA satellite (vegetation monitoring) were in a rare geographic projection, the Hammer Aitoff projection, which was hardly supported by any of the than more popular GIS software programs.

The task at hand for the RRSU was to start a number of activities to develop standards for the digital databases and the objective was to develop a standard raster and vector database for the SADC region, which would allow easy use and analytical procedures in a GIS environment and facilitate regular updates.

Organisational Approach

Overall Leadership - The SADC RRSU provided overall leadership for this regional activity. The RRSU identified needs and formulated the plan; implemented development with strategic partners; assessed availability of data; organised the data collection; ensured evaluation and quality control of the outputs; and distributed the output.

Development was accomplished by the SADC RRSU. Technical partners in the development were the Office of Arid Land Studies of the University of Arizona, and the University of Stellenbosch. Both Universities were responsible for technical tasks, which were implemented under a contractual agreement. Development of the digital spatial databases involved the processing of data, creation of basic data layers, preparation of documentation, and the development of the system on transportable media with a user-interface to view and analyse the data.. As a starting point, several layers of the Digital Chart of the World (DCW) were used, as well as the Africa Data Sampler (ADS) prepared by the World Resources Institute (WRI, Washington - USA). The WRI provided the RRSU with a pre-release of the ADS in 1994 in order to facilitate a first review of the available data. The internationally available data was merged with existing national digital data sets. Where necessary, hard-copy maps were digitised, corrected and georeferenced. This was done by the University of Arizona, while at a later stage the University of Stellenbosch was contracted to review and correct the soil database.

The RRSU was responsible for the processing of all satellite image raster data into a 6-minute geographic projection. Using this standard format, data from different satellites, or the same satellite, but received by different data acquisition systems, are in the same geographic format and can be used together with the vector data in a wide range of GIS applications.

Since 1994, the development has gone through several phases and has resulted in a uniform and standard satellite (Meteosat and NOAA) image database; a standard and uniform thematic vector database at the scale of 1:1million. A first version of the vector database was released on CD in 1995. In June 1997 the first version of the "RRSU CD" was released, which included also all satellite data, agricultural statistics, and basic climate information. An update was released in March 1998. The RRSU CD also includes a software facility to view and analyse the data, called "WinDisp". This program was developed with financial support of a number of partners, including the RRSU. A next release is expected in the first half of 2000. More recently, in June 1999, the RRSU has produced a second CD with a detailed regional climate database in raster and tabular format, including information on rainfall, temperature and evapotranspiration.

In addition to this, the Harare based: Aquatic Resource Management for Local Community Development Programme (ALCOM), used the hydrology layers from the RRSU spatial database to develop a comprehensive hydrological data base and watershed map for Southern Africa which is fully compatible with the standard format established by the RRSU.

Other major stakeholders in the development phase included: (i)The National Early Warning Units (NEWU's); and (ii) the National Meteorological Departments (NMD's) in the SADC countries who played an important role in evaluating the data sets and provided suggestions for corrections or better data. Other major data contributors included organisations such as: (i) the World Resources Institute; (ii) USGS Eros Data Center; (iii) FAO; (iv) UNEP GRID; and (v) the USAID Famine and Early Warning System (FEWS). Regional or national level data was provided by: (i) the NEWU's; (ii) NMD's; (iii) National Remote Sensing Centre's; (iv) Environmental Councils; and (v) various Government Departments in the SADC member states. User beneficiaries include Government institutes; Ministries; national, regional and international organisations; private trading and industrial sector; banking and finance groups; large-scale and small-scale farming organisations; and NGO's.

Review and evaluation of the effort for meeting the needs of SADC members was performed by the SADC RRSU; the National Early Warning Units and National Meteorological Departments in the SADC countries. The review and evaluation process included making data available for evaluation; conducting workshops/meetings to introduce the databases; collection of evaluation comments/reports; and ensuring incorporation of corrections/additions.

Distribution of the database, tools, meta-data, and viewing and analysing software was accomplished by the RRSU. The RRSU make data available in a user-friendly format on CD, sponsor workshops/meetings and maintain an internet web-site to create and maintain awareness, encourage and act on suggestions and recommendations, and are responsible for regular updates of the data bases. The new historical database was distributed to all contact points in the SADC Member States. Backstopping missions and regional workshops were used to inform the contact points about the changes and the characteristics of the new data format.

Traditionally, Internet accessibility in Africa has been significantly low compared to other regions of the world. Although Internet accessibility is improving rapidly in the SADC region, the RRSU will continue to distribute the data on CD. The reason for this is that: (i) the size of the RRSU spatial data sets are to big to be used operationally over the Internet (even with high speed access), and (ii) using the data structure on the RRSU CD and the include software, the data can be viewed and analysed. However, at present the RRSU is improving its local Internet connectivity through the installation of a radio-link to one of the major Internet Service Providers (ISP) in Harare. With this installation in place FANR (and the RRSU in particular) will have the possibility to offer their data bases on-line over the Internet using their own server capability. However, it should be noted that even when data is offered over the Internet: (i) many stakeholders will still have limited access; and (ii) the specific analytical capability offered on the RRSU CD will not be available.

Users include many of the stakeholders noted above, which include the National Early Warning Units and National Meteorological Departments in the SADC countries. A range of government institutes; Ministries; national, regional and international organisations; private trading and industrial sector; banking and finance groups; large-scale and small-scale farming organisations; and NGO's use the system as well.

Finally, although the RRSU used contractual agreements with the University of Arizona and the University of Stellenbosch for development, collaboration with other partners was basically established through informal agreements. Data was normally provided as part of a mutual agreement, in that the RRSU would correct/update the data sets and return it in the new format to the data providers.

Programme Successes and Issues

The success is obvious. The RRSU databases provided on CD are in high demand. The capability is considered by many to be the regional standard and even in many SADC countries it is considered to be the best and most complete data set available. However, there is no formal regional SDI structure for the SADC region, though informal initiatives are undertaken to reach consensus. A good example is the collaboration between SADC RRSU and the National Spatial Infrastructure Framework (NSIF) in Pretoria - South Africa. Together with a number of other stakeholders in SADC, and the remainder of Africa, a number of activities are being launched to formalise a SDI policy body.

Implementation Approach

The RRSU has introduced a regional standard for spatial data, which is now being adopted in a number of SADC countries. This data standard has been presented during different meetings. An example is the SADC Environmental Information Systems (EIS) network. During a meeting of representatives of the SADC EIS Network in November 1997 a number of very general recommendations were made about the scale and format of spatial data. The RRSU spatial database was used as an example. However, at the same meeting it was agreed that this format should be used as a common data "exchange" format and that it is up to the countries to decide what format is used at national level.

On behalf of the SADC EIS Network, the SADC Environment and Land Management System (ELMS) has been working on a data policy document, which will be available by early 2000.

In conclusion, the RRSU database development essentially introduced regional standards, which are now being adopted by SADC member countries. Although the standardisation efforts were driven largely by the need to establish viable RRSU databases for early warning for food security, is clear that there are more potential applications of the data against different issues (such as natural resources management). The RRSU spatial databases were prepared in response to critical and specific needs for the SADC member nations with regards to early warning for food security. Without a clearly defined and consistent SDI nationally and regionally, the RRSU worked with members and stakeholders to establish the core infrastructure components needed to accomplish development and implementation objectives. The following is a chronology of events and actions completed to complete the Early Warning effort:

In 1994, the RRSU began work with stakeholders to assess the need for uniform data standards for the SADC region, and to identify the partners needed to accomplish the development. This included the preparation of contracts in some cases.

- In 1995 development focused on the collection of data for the vector database. As noted above, data came from a number of sources, with data provided compliant with international standards, along with other non-compliant data that needed to be processed to an acceptable standard format (implemented by the University of Arizona). Database development and evaluation also occurred at this time, including the review and correction of the SADC Soils database (implemented by the University of Stellenbosch). At a regional workshop in September 1995, new data standards for raster data were introduced and accepted.
- Throughout 1996, database information was distributed and reviewed by member nations. Due to the lack of regionally consistent data standards and formats, data had to be converted to the native format of member countries for review. Evaluation results were reviewed and documented. From June to December 1996, the transfer of IDA analytical functions to the application software WinDisp (financed by the RRSU and implemented by the University of Arizona) was accomplished.

Based on evaluations provided by stakeholders, changes were made to the vector database in early 1997. A user-friendly interface was developed for the user application, and other structure and file naming issues were resolved. Member nations each received a pre production CD for review during this period. By summer of 1997, the completion of the CD was announced, and distribution commenced.

- By early 1998, RRSU had issued an updated version of the Early Warning system, and had begun routine maintenance and update of the data sets to ensure information utility for the region. In conjunction with South Africa, RRSU commenced training on metadata creation and implementation.

The RRSU spatial data base program has been of major benefit to the SADC region. With agriculture recognised by member nations as a major area of mutual interest, the SADC now promotes regional cooperation and economic development through a Program of Action covering cooperation in various sectors. These sectors include those related to food, agriculture and natural resources. Food security and natural resources management is one of the main pillars for economic development and social welfare in the region.

A solid, harmonised and uniform regional spatial database contributes to an improved information in support of managing scarce resources, which are required to secure food security and human well being in the region. In addition, the FAO Global Information and Early Warning System (GIEWS) are using the data from the RRSU spatial databases. Moreover, the GIEWS Internet web site links directly into the SADC FANR Web-site, a good example of sharing information and not duplicating it!

Conclusions

The RRSU database activity has helped focus the SADC region on establishing the basic elements of a national and regional SDI. However, further progress toward a healthy and responsive regional SDI will depend on the resolution of a number of important issues. Several of the major issues facing the region are summarised below:

Telecommunications Infrastructure - Although the initial RRSU spatial database program has focused on establishing standards for data exchange, efforts are underway to establish improved dissemination capabilities via the Internet. However, until the telecommunications infrastructure is more available to stakeholder organisations, SDI delivery will be limited to physical products, information and services such as the CD-ROM based applications and data associated with the RRSU program. However, it should be note that the RRSU spatial database a rather big in size and in order to work with the data on an operational basis the CD-ROM will be the most applicable medium for distribution.

National and Regional SDI Policy - From and organisational and policy point of view, formal SDI policies and practices as the national and regional level are still forming. At this stage there is a need to create higher level of awareness of the benefits of a compatible SDI for the region and its nations. Furthermore, there should be a formal review or survey of the specific state of each member nation in terms of SDI development or plans. The RRSU took every opportunity to demonstrate the need for a uniform SADC database. And, much of the RRSU's success has been accomplished through informal contacts, which have contributed to the process of awareness and willingness to share critically important data sets to this regional initiative.

Data ownership and pricing policy - There are still unresolved issues regarding data ownership and pricing policies. This has been particularly true with climate data. The NMD's in the SADC region are following the advise of the World Meteorological Organisation (WMO) that climate data should be made available on a commercial basis. Since the NMD's are SADC institutes they have made data available to the RRSU in order to develop a regional tabular database and create climate (raster) layers to be used for analytical purposes and research. The RRSU is not in a position to distribute these tabular data sets or climate layers. What will be done is that the RRSU will train the NMD's in the concept of creating these databases and data layers. The NMD's will than be able to distribute the databases.

Recommendations

Education and Awareness – Establish a clear program of education and awareness building to gain support of national policy makers across the region. This program should include the assessment of each member nation, and the identification of issues and areas of focus to establish compatible SDI's that address both national and regional issues

Organisation and Partnerships – further work needs to be accomplished in getting a basic and flexible structure for SDI development at both the national and regional levels. Formation of a more formal SDI Committee for the African continent with appropriate regional sub elements may help further organise and encourage collaboration

Funding – Long-term commitment of funding must be obtained to develop, implement, and maintain a regional SDI on a continuing basis. While external funding sources have resulted in measured success in the SADC region to meet specific objectives, pervasive funding from both internal and external sources must be secured to assure that a compatible SDI is created for the region. One major lesson learned through the RRSU program is that funding for data maintenance must be included in SDI operations to assure that spatial information remains relevant to decision makers.

Standards – Member nations of the SADC must continue to identify standards that create compatibility for data content and metadata throughout the region. Regional standards should be based where possible on existing international standards, and when new standards are needed, SADC members should participate where possible in the formalisation of standards at the international level when appropriate.

Telecommunications – The lack of Internet access among member nations continues to be a major issue for the region. Continued focus on expansion of Internet services and increased access by member nation users of spatial information and services must be supported. Because the improvement of Internet access in the region will take some time to develop, the availability and distribution of data, as well as meta-data, should be done using other sources as well. Therefore the distribution of this type of information on CD-ROM, using the latest digital technology should be considered.

Policies on data ownership and licensing - There is a need for a clear data policy in the region which include sections on intellectual property rights, distribution mechanism and pricing of data. This should be addressed not only within the SADC region, but also as a major initiative of the GSDI to achieve a greater understanding of the international and global implications of data ownership, licensing, and usage.

Global Case Studies – Activities Contributing to a Global Spatial Data Infrastructure

Mindful of the critical social, environmental, and economic issues shared regionally and often globally, the assurance of a Global Spatial Data Infrastructure to enable cooperating nations and organisations to collaborate on issues and solutions is extremely important. Without a global reference environment where a consistent set of policies, standards, best practices and co-operating organisations guide national and regional spatial data infrastructure development, we run the risk of being unable to effectively and jointly address pressing issues in the global context.

Today, there are a number of major initiatives that address one or more of the components of the Global Spatial Data Infrastructure as defined by the GSDI Committee in March 1999. Indeed, the GSDI's success is dependant on the successes and compatibility that many of these programs bring to the global marketplace – technology, data, standards, resources, organisational mission, and distribution. This section of outlines some of the major the contributors toward a GSDI. The Digital Earth Initiative, launched in the United States, China and other nations is reviewed as an example of a program that has the potential to focus and accelerate research and development programs needed to achieve the vision of a Digital Earth (www.digitalearth.gov) and the critical supporting infrastructures needed at the local, national, and global levels. Finally, this section includes a discussion of remaining areas of challenge toward the formation of a pervasive GSDI.

GSDI Defined

At the 2nd GSDI Conference in 1997, the multi-national GSDI Steering Group defined the Global Spatial Data Infrastructure as:

"… The policies, organisational remits, data, technologies, standards, delivery mechanisms, and financial and human resources necessary to ensure that those working at the global and regional scale are not impeded in meeting their objectives..."

An Overview of GSDI Infrastructure Elements

Given this definition, it is important to note that a number of programs address various aspects of the GSDI at a global level. This section summarises some of the major programs that have contributed to a Global Spatial Data Infrastructure. This list is by no means exhaustive, and in fact has been abbreviated to provide examples of the work that is being accomplished towards a GSDI.

For example, the International Steering Committee for Global Mapping is working to produce a Global Map, to be released in 2000. The United Nations has had in place since the 1980s a Global Resource Inventory Database and other similar resources. The International Geosphere Biosphere Programme is working to provide global environmental data sets to scientists. The Open GIS Consortium (www.opengis.org) is working to promote technological and computing advances that can support the development and use of environmental data and their accompanying infrastructures. The International Standards Organisation Technical Committee 211 (http://www.statkart.no/isotc211/welcome.html) is developing a metadata standard.

The International Steering Committee for Global Mapping (ISCGM) (http://www1.gsi-mc.go.jp/iscgm-sec/index.html) was created as a response to Agenda 21 from the United Nations Conference on Environment and Development (UNCED) held in Rio de Janeiro in 1992. Chapter 40 of Agenda 21 was a call for global environmental data. As a result, the Japanese Geographical Survey Institute/Ministry of Construction took the lead on the project and formed the ISCGM in 1994. Membership in ISCGM is comprised of representatives from national mapping agencies, nongovernmental agencies, and academia. The result is a project involving sixty-five different national mapping agencies and other organisations from every continent on the earth. The goal is production of Global Map, which will contain elevation, vegetation land use, drainage systems, transportation networks, and administrative boundaries, all at the nominal scale of 1:1,000,000. In the process, focus on a strategic plan, specifications, and data policy has been necessary.

In addition to UNCED, the United Nations has other organisations that play a role in the creation and dissemination of environmental data. Often, these organisations have mandates to create and make these data available. The primary environmental data organisation of the UN that comes to mind is the United Nations Environment Programme (UNEP) Global Resource Inventory Database (GRID) (www.grid.unep.org). GRID was formed "to assist UNEP and its partners by contributing environmental data and information, as well as methodological techniques for handling such data, to enhance the scientific basis for decision making and help advance sustainable development initiatives." GRID is a network of sites located around the world, all of which provide environmental data. UNEP/GRID is composed of a variety of sites (Arendal, Norway; Bangkok, Thailand; Christchurch, New Zealand; Denmark; Geneva, Switzerland; Kathmandu, Nepal; Moscow, Russia; Nairobi, Kenya (headquarters); Ottawa, Canada; Sao Jose dos Campos, Brazil; Sioux Falls, USA; Tsukuba, Japan; Warsaw, Poland). Each site provides some global data sets, but most often, they have a specific focus. For example, the Kathmandu site focuses primarily on mountain related issues and data.

In addition to UNEP/GRID, the United Nations Educational Scientific and Cultural Organisation (www.unesco.org) has played a role in the development of global soils databases. In addition to UNESCO, the UN Food and Agriculture Organisation (FAO) (www.fao.org) played a leading role in the development of the 1:5,000,000 global soils database in the 1970s. FAO also has several programs within its jurisdiction, including the Global Information and Early Warning System, which "monitors the crop and food outlook at global and national levels to detect emerging food shortages and assess possible emergency food requirements." The FAO's Forest Resources Assessment (FRA) is a decadal tree census, and is used to help determine rates of deforestation. The United Nations Development Programme (UNDP) (www.undp.org) also has an interest in global data set development efforts and has supported research in this direction.

The International Geosphere Biosphere Programme (IGBP) is a programme within the International Council of Scientific Unions (ICSU). Within the IGBP is the Data and Information System (IGBP-DIS) (http://www.cnrm.meteo.fr:8000/igbp/index.html). The goals of IGBP-DIS are "to describe and understand the interactive physical, chemical and biological processes that regulate the total Earth system, the unique environment that it provides for life, the changes that are occurring in this system, and the manner in which they are influenced by human actions."

IGBP research currently focuses on six key questions that are addressed by eight Core Projects:

How is the chemistry of the global atmosphere regulated and what is the role of biological processes in producing and consuming trace gases?
How will global changes affect terrestrial ecosystems?
How does vegetation interact with physical processes of the hydrological cycle?
How will changes in land-use, sea level and climate alter coastal ecosystems, and what are the wider consequences?
How do ocean biogeochemical processes influence and respond to climate change?
What significant climate and environmental changes have occurred in the past and what were their causes?

Three crosscutting Framework Activities that include assists the integration of IGBP Core Projects:

IGBP Data and Information System (IGBP-DIS)
Global Analysis, Interpretation and Modelling (GAIM)

Global Change System for Analysis, Research and Training (START), addressing regional research initiatives and needs, jointly with the IHDP and WCRP.

Examples of the data available through these efforts include the global land 1 km AVHRR data set, the IGBP DISCover data set developed from the AVHRR data, as well as the global FIRE data.

The OpenGIS Consortium (http://www.opengis.org/) is an organisation "whose mission is to promote the development and use of advanced open systems standards and techniques in the area of geoprocessing and related information technologies."

The International Standards Organisation Technical Committee 211 (ISO/TC211) (http://www.statkart.no/isotc211/welcome.html) goal is "standardisation in the field of digital geographic information." According to their web site:

- This work aims to establish a structured set of standards for information concerning objects or phenomena that are directly or indirectly associated with a location relative to the Earth.
- These standards may specify, for geographic information, methods, tools and services for data management (including definition and description), acquiring, processing, analysing, accessing, presenting and transferring such data in digital/electronic form between different users, systems and locations.
- The work shall link to appropriate standards for information technology and data where possible, and provide a framework for the development of sector-specific applications using geographic data.

The organisations and activities shown here do not cover all the activities described in the Global Spatial Data Infrastructure definition. ISCGM is focusing on the data, standards, and organisational commitments to generate and maintain a global framework of key geodata themes. The Open GIS Consortium is interesting in promoting technological advancements and standards. The ISO/TC211 is aiming toward the standardisation of environmental metadata. And the Digital Earth Initiative (discussed in detail below) is working to link together many of these activities to focus research, development and partnerships necessary to advance capabilities needed to sustain the Digital Earth vision. Together, these different, and seemingly disparate, activities can create a greater whole that can benefit many different people and organisations.

A cube illustrates the contributions and relationships of the various organisations around the world that have helped shape the GSDI. National and regional SDI efforts represented on the one side of the cube illustrate the major resources, technology, metadata / data standards, and best practices shared internationally. Many of the standards, technologies and practices have been adopted or have influenced international standards are shown on a second face of the cube. On the third face of the cube are organisations and activities, which have contributed to specific areas of the GSDI. FAO / GRID have produced global soils data, the Global Map aims to provide a consistent global set of geographic coverages, along with the commitment of nations to maintain the data. The Open GIS Consortium and International Standards Organisation bring data and metadata standards to the global community for use by all nations and organisations.

Indeed, the efforts of these organisations have yielded key elements of the GSDI, many of which have become part of the overall GSDI reference environment needed to help gain compatibility at a transnational and global level. However, much more work needs to be accomplished to address the remaining technology, policy, and resource issues that are limiting the implementation of the GSDI. The Digital Earth initiative is discussed below as one example of an activity focused on addressing some of the major challenge areas related to GSDI.

The Digital Earth - a Case Study in the Genesis of a Global Spatial Data Infrastructure

In 1998, United States Vice President Al Gore communicated a vision for the future and the way citizens would interact with global information resources to better comprehend the complexity of our planet and our interactions with it.

A United States Digital Earth Interagency Working Group developed consensus that the Digital Earth Initiative involves a national and international effort to plan and build a cooperative use, Internet-based infrastructure to use vast quantities of geo-referenced data and information resources, Earth science data, and cultural and historic data. This query based and visually oriented data will be used by federal, state, local and tribal government communities, academia, and the private sector for scientific applications, practical decision-making, education, journalism, and other citizen accessible applications. As user interface prototypes become available, it will also be possible to interact with Digital Earth through Internet portals around the country, and obtain a better level of access and interoperability with the Earth's geospatial, social, and economic data (www.digitalearth.gov).

Success of Digital Earth is directly correlated to the soundness of the infrastructure it uses as a foundation. In addition, myriad protocols and standards arriving with the World Wide Web must be accounted for in the development process. Network infrastructure for Digital Earth will be based on the U.S. National Spatial Data Infrastructure (NSDI) and the Global Spatial Data Infrastructure (GSDI). Leveraging of these programs is required to ensure full utilisation of best practice for creation of a core infrastructure.

One of the major challenges for Digital Earth is to construct the organisational structure that will enable citizens, industry, academia, and government interaction in developing the initiative. These communities must then coordinate the focus of research and development requirements to create the Digital Earth. Identification of technology, organisational, policy, and other barriers to success needs to be well articulated among the various organizing bodies to better implement solutions. The Digital Earth initiative will work to focus the resources of its partner organisations to accelerate solutions to barriers that prevent or limit the achievement of the Digital Earth vision.

The Digital Earth must also achieve a strong public-private partnership to link industry and other non-government organisations with government. Government agents must continue to conduct policy and technical meetings to support the PPP and the international community. At present, the U.S. has a federal government structure in place and is working with industry, non-governmental organisations (NGOs), and academia to nurture a sustaining membership for the PPP. At the international level, the Chinese have instituted the international Digital Earth symposium (the first held in Beijing, December 1999, with 25 countries) to be held biannually.

A characteristic of Digital Earth for outreach and education is the public engagement value through the application of impressive 3D visualisation and immersive-interactive computer technology display stations. Museums have experienced much success in capturing the public's attention with Digital Earth displays that provide global perspectives of the planet using satellite monitoring technology. As the demonstration, test beds, and scenarios increase the Digital Earth content, the public, including industry and education, can be expected to increase awareness and support of this initiative. This enhances support of the cross cutting program, that is GSDI and NSDI, which have less connectivity with the popular media.

- Development of a strategic plan with a support community is requisite. A useful scheme for defining the major components, or development areas for the Digital Earth Initiative helps in focusing resources where they are most needed. Six development areas have been identified as follows:

- Visualisation and Exploration (focused on the methods, hardware, and software for viewing and exploring Digital Earth data; involves the user community through the information science and human factors researchers and Information Technology companies);

- Education and Outreach (focused on the users, scenarios, and partnerships that add value and relevance to the DE; involves the user community through museums, schools and the media);

- Science and Applications (focused on the development and validation community for Digital Earth content; involves the user community through scientists, state and local governments, and commercial application developers);

- Advanced Display Sites (focused on the projects, test bed prototypes, and facilities through which the Digital Earth gets tested and used; involves the user community, such as NASA centers and museums);

- Data Access and Distribution (focused on the gathers and distributors of geo-referenced data; involves the user community through network bandwidth providers and Earth Science Federations (e.g., DAACs));

- Standards and Architecture (focused on the infrastructure and interoperability protocols for a sustainable Digital Earth; involves the user community through organisations such as CEOS, OGC, FGDC, and NMOs).

Digital Earth is dependent upon many factors in the technology fields that may cross cut through any one of the six development areas. Assessments of the technology challenge will remain a consistent part of the Digital Earth initiatives so that as technology gaps are identified; resources can then be marshalled to address these gaps. Coordination with the National Academies of Sciences must be maintained to conduct assessment in computer technology, web networks, advanced algorithms, remote sensing, as well as the mapping sciences. The following technology development areas have been highlighted for the Digital Earth Initiative:

- Computational Science (e.g., high-speed computing for modelling and simulations; integration and overlaying of diverse sources of geo-referenced information, interactive 3-D visualisation, display and navigation, computation of information products on demand);

- Mass Storage (e.g., distributed active archives with real-time access of large, multi-resolution data sets);

- Satellite Imagery (e.g., 1 meter to one kilometre seamless resolution for the planet);

- Broadband Networks (e.g., high-speed networks and public access nodes for transmission, interaction, and collaboration);

- Interoperability (e.g., Internet and World-Wide-Web standard protocols); and

- Metadata (e.g., advances in automated database documentation software).

The success of the Digital Earth Initiative is heavily dependent on the continued progress of national, regional and the global SDI initiatives and other global geospatial programs discussed in this Cookbook. The impacts of policies, technologies, and organisations at local, national, and international scales are interdependent and therefore complex. Digital Earth provides an overarching vision for the future that may well benefit the creation and maturation of the GSDI and associated programs through the collaboration of efforts for these challenging developments.

More information on the Digital Earth Initiative can be found at www.digitalearth.gov. A draft version of the Digital Earth Reference Model (DERM) can be found at www.digitalearth.gov/derm/.

Summary - Furthering the Global Spatial Data Infrastructure

The case studies and recommendations in this chapter, along with the information provided elsewhere this document have detailed the many initiatives underway that are contributing towards the objectives of the GSDI. However, much more work needs to be accomplished if the GSDI is truly to be a global resource from which all nations and organisations can access resources to build compatible infrastructures. Further advancements in data, standards, delivery, and technology are needed. However, a much more focused effort needs to be placed on outreach and education, resources, policy and legal issues related to SDI development if GSDI objectives are to be achieved.

In responding to these needs, the GSDI Steering Group has initiated a number of initiatives in calendar year 2000 to further advance the objectives of the GSDI:

Business Case Study - Emphasis is being placed on the development of a Business Case for Spatial Data Infrastructures. The study will identify the economic, social, environmental, and disaster management benefits that can be achieved through development of compatible national and regional SDI's, and the global SDI.

Address Legal and Economic Issues– The GSDI Steering Group has formed a Legal and Economic working group to focus on addressing the implications and potential solutions to legal and economic (funding) mechanisms that underpin the GSDI

Improve Outreach and Communications – the Communication and Awareness Working Group will focus on developing and implementing the programs necessary to raise awareness, articulate the value and secure additional support for the GSDI.

Your support of the Committee and working groups is encouraged. Nations must be able to establish Spatial Data Infrastructures that address internal matters of concern, while providing the ability to work at the transnational and global levels to address the important issues such as those outlined by the UN Agenda 21, the Kyoto Protocol. Please contact us at www.GSDI.org, and help us achieve our goals. Together, we can establish and SDI that allows us all to act locally, nationally, and globally.

Chapter Ten: Terminology

Editor: Antony Cooper, South Africa

Note to Readers: Comprehensive terminology section will be prepared for Version 2.0.