GINIE: Geographic Information Network in Europe

IST-2000-29493

Registries and e-Services:

Final Report

D 2.6.2 (b)

Project Co-ordinator:

University of Sheffield - USFD

Partners:

Open GIS Consortium (Europe) - OGCE

European Umbrella Organisation for Geographic Information - EUROGI

Joint Research Centre of the European Commission – JRC

Registries and e-Services:

Final Report

D 2.6.2 (b)

Editors:
Lance McKee
Günther Pichler

CONTENTS

1. Introduction

Sharing geographic information and services among different user communities is the main challenge facing producers and users of geographic information. This report details the recommendations that came out of GINIE's workshop on Registries and e-Services (Munich 21-23 January 2003). Workshop attendees focussed on the technical and policy issues involved in implementing online registries that facilitate discovery of and access to Web-accessible geospatial data and geoprocessing "e-services".

Two other GINIE documents, *Towards Geospatial Semantic Web Services* and *Catalogue Mini-Cookbook* are included as appendices to this report. They provide additional technical information related to these issues.

The overall recommendation of this report is to move from our current mode of a few people having access to a few static geospatial data repositories and stand-alone GIS systems to many people having real time network access to many dynamic geospatial data repositories and online geoprocessing resources. The vision is geospatial information fully integrated into the information society. This is both a cultural and technical challenge.

1.1 Registries' Role in Spatial Data Infrastructure (SDI)

There is an increasing recognition that many of the main challenges of modern society such as environmental protection, security, transportation, socially just development and enhanced services to citizens require decision-makers to have easy access to spatial (i.e. geographic) information. Such information must not only exist and be kept current, it must also be easy to discover, evaluate, access and integrate with other information. This requires a framework of policies, institutional arrangements, technologies, data and people that makes it possible to share and use geographic information. The term **Spatial Data Infrastructure (SDI)** encapsulates such a framework. The importance of an SDI for good governance and for economic and social development has led most countries to begin to develop such infrastructures. In Europe, most countries are in the process of developing SDIs at national and/or regional/local levels. (See the GINIE report, *Spatial Data Infrastructure: country reports*; http://www.ec-gis.org/ginie/documents.html.)

The development of the "Information Society" has been one of the major policy goals of the European Commission since the Maastricht Treaty in 1993 gave the Union the responsibility of supporting the integration of trans-European networks in transport, energy, and telecommunications. INSPIRE (INfrastructure for SPatial InfoRmation in Europe: http://www.ec-gis.org/inspire) is an important part of the information infrastructure that supports these physical infrastructure networks. The overarching policy principles of INSPIRE are:

1. Data should be collected once and maintained at the level where this can be done most effectively;
2. It must be possible to seamlessly combine spatial information from different sources across Europe and share it between many users and applications;

3. It must be possible for information collected at one level to be shared between all the different levels, e.g. detailed for detailed investigations, general for strategic purposes;
4. Geographic information needed for good governance at all levels should be abundant and widely available under conditions that do not inhibit its extensive use; and
5. It must be easy to discover which geographic information is available, fits the needs for a particular use and under what conditions it can be acquired and used.

There has been progress in addressing this set of requirements, as described below.

A GINIE expert meeting, the GI Data Policy Workshop, was held 23-24 May 2002 to identify areas of similarity and difference in the GI data policies of selected European countries and to discuss possible policy options for the proposed legal instrument implementing INSPIRE. The INSPIRE Working Group on Data Policy and Legal Framework was well represented in this meeting.

Seven months later, the 21-23 January 2003 workshop in Munich on Registries and e-Services began with a conclusion of that May 2002 Data Policy Workshop: *In the past, geographic information catalogues and gazetteers were in the form of books, but in the future they will be computer-based systems.* The Registries and e-Services workshop focussed on the IT components and the data methodologies necessary for the development of an SDI. It addressed semantic interoperability, with a focus on technologies (registries and e-services) and practices that could be implemented in Europe to support efficient discovery and sharing of spatial data. The workshop drew 23 experts from 9 European countries and the US. The list of participants is included in Appendix C.

This report details the findings of that meeting focusing on the practices, technologies and the political support needed to develop e-registries. It discusses several key areas that are useful to define.

e-Services are network-resident processing services. They have software interfaces that enable diverse applications with conforming interfaces to reach across the network and programmatically "bind" or couple with the services, almost as if they were subroutines residing on the same computer with the application. (A software interface is a shared boundary between two programs, defined by an ordered set of expected parameters and corresponding functions set to read those parameters.) Both registries and GeoPortals (see below) depend on e-services. E-services on the Web are commonly called Web Services.

Computing that works across a network via e-services is commonly called "**distributed computing**.""

Metadata is data about data or data about e-services. Metadata that describes spatial data includes detailed information about when a set of this data was collected, how it was collected and what naming schema was used in naming geographic features. Service metadata includes detailed information about what processing an e-service can provide and information about the details of the interface by which the service can be invoked and used.

Registries are essentially the same as **catalogues**, for the purposes of this discussion. They help users or application software find and use data or e-services existing anywhere in a distributed computing environment.

A **GeoPortal** is an assembly of components that provides a community-wide, Web-based access point to distributed data and processing resources, including, for example, registries. A portal often serves a specific community, but it may use a generic user interface that other communities can adapt. A portal usually offers personalised or customised views of some kind.

Political support for SDI development is necessary at the national, regional and local levels. SDI policy issues include: access rules, charging and licensing, data quality, delivery methods and, most importantly, from the standpoint of registries and e-services, support for standards that enable technical and semantic interoperability.

1.2 Interoperability is essential.

The key to information sharing is **interoperability**. Interoperability is the ability of autonomous, heterogeneous, distributed digital entities (e.g. systems, applications, procedures, registries, services or data sets) to communicate and interact or be used together despite their differences. Two types of interoperability are necessary for efficient sharing of spatial information:

- **Technical interoperability** refers to the ability of different geoprocessing software systems to communicate and interact through shared interfaces; and
- **Semantic interoperability** refers to standards that support the ability of people and software systems to find and use spatial data produced at different times by different people for different purposes, in which geographic features may be represented using different naming schemas and "geometries". ("Naming schema" refers to a "data dictionary" of geographic feature names. "Geometry" refers to spatial reference systems for representing the location of features). Semantic interoperability depends largely on data producers adhering to standard feature naming schemas and metadata schemas. Some of the difficulties associated with this will ease when "semantic translation" becomes technically feasible.

This report addresses both technical interoperability and semantic interoperability issues related to registries and e-services.

2. Registries and e-Services

Like a card catalogue or computer index in a conventional library of books, a registry reduces the number of places the user or application has to look when searching for a particular resource. Digital registries and e-services are fundamental components of the European Spatial Data Infrastructure (ESDI) because the ESDI includes large numbers of distributed and already existing digital data sets and computing services, far too many for anyone to know about. It makes sense to organise them in a network for automated access by authorized users. The Web, conveniently, provides the basic network infrastructure.

Online registries and e-services are components of distributed computing, in which software running on nodes of a network is able to use software and data on other nodes of the network.

2.1 E-Services

E-Services are network-resident processing services. They are accessible through interfaces that enable diverse applications with conforming interfaces to programmatically “bind” or couple with them. That is, an application can send instructions across the network to instruct an e-service on a remote server to perform a processing task, and the e-service will return the result of the processing back to the application. E-services can invoke other e-services, which is called “service chaining”.

E-services called “Web Services” are the basis for today's dominant distributed computing paradigm that based on the Internet and the Web. The Web Services architecture is, like the Web that supports it, based on open interfaces. In open Web-based distributed computing, registries make it possible for online services to be found and used. (In a more closed Web-based distributed computing scenario, Application Service Providers, or ASPs, provide online services, but the service links are small in number and fixed rather than being numerous and available through a catalogue. Such control, of course, is an advantage in some cases.)

An example of a spatial e-service is the use of a remote coordinate transformation service. In Web Mapping based on OGC's OpenGIS® Web Map Service Specification:

1. An application instructs a remote Web Map Server to provide a simple raster image (a GIF, perhaps) of a specified rectangular region of a specified data layer. The raster map that is returned for display by the client presents the data in a particular coordinate reference system;
2. The application may then instruct a different Web Map Server to provide a simple raster image of the same geographic region of a different data layer, for display and overlay by the application. Of course, the data in the second image must be presented using the same coordinate reference system as the first image, or else the two cannot overlay accurately; and
3. If the second Web Map Server cannot perform the necessary coordinate reference system transformation, that server or the application can send a request to an e-services server that provides a coordinate transformation service. That server receives the data and the conversion instructions from the application, performs a coordinate transformation on the data, and returns the transformed data to the data server, where it converted to a raster image and sent to the application for display.

2.2 Registries

A registry is a set of special e-services that support organisation and discovery of and access to online data and processing services (e-services). Registries help users or application software or other services find or retrieve data or e-services existing anywhere in the distributed computing environment.

Registries are used in the larger Web Services world, but in this report they are discussed and described in the ESDI context of geodata and spatial technologies. In this context, registries are catalogues of both online geodata and online geoprocessing services. Building on the Web Services architecture, the Open GIS Consortium (OGC) has built an open set of special Web Services for geoprocessing, called “OGC Web Services”.

Registries, like card catalogues in a library, store and enable searches of descriptive information about the resources described, but do not store the resources themselves. A registry can be thought as a specialised database of information (metadata) about data and e-services. Owners of network servers that serve data or e-services register metadata about these resources in a registry so the resources are published resources. Special registry e-Services (Catalogue Services, in OGC's terminology) provide:

- Methods for such registration of metadata and methods for changing the metadata;
- Methods for searching the metadata collections to discover and evaluate resources whose metadata is published in the registry; and
- The method by which the application connects to the online data source or spatial e-service.

OGC's Catalogue Services Specification provides a framework that enables multiple registries to be “viewed” as one registry. Stewards of geodata and online geoprocessing services in a nation, a discipline, a metropolitan area or an enterprise can virtually combine, or “federate”, their registries if their registries conform to the OpenGIS Catalogue Services Specification. A query to one can thus query any or all of the others. This multiplies the potential value of the data and services whose metadata are published in any of the federated catalogues.

Much spatial data at the present time, of course, is off-line, but metadata for such data can nevertheless be registered in a registry. This situation describes today's spatial data **clearinghouses**. A clearinghouse helps a user find and evaluate data, but it does not usually provide the means for automatically viewing, downloading or operating on the data. Most clearinghouses around the world are being updated to conform to OGC's OpenGIS Catalogue Services Specification. Countries implementing registries based on OGC's OpenGIS Catalogue Services Specification include Australia (CANRI's NSW Natural Resources Data Directory), Canada (Geoconnections Discovery Portal), the UK (AGI's GI Gateway), Germany (GDI NRW), Netherlands (RAVI's NCGI), South Africa (National Spatial Information Framework), Spain (IDEC in Catalunya) and others.

Registries enable:

- Resources to be organised in various ways without changing the content or physical organisation of the data;
- Users to find information about resources;
- Access to desired resources once they have been located; and
- Multiple ways of viewing the same resources.

Registries also present a common format for the metadata describing each resource, but this presupposes that the metadata provided for registration in the registry was provided in a standard format.

A more detailed introduction to building registries for e-services can be found in the "mini-cookbook" entitled *How to Build Catalogues for e-Services*, by OGCE/OGC.

2.3 Metadata and XML

The UK Office of the e-Envoy has published an "e-Government Metadata Standard (e-GMS)" Version 2.0, dated 16 May 2003, which begins with the following list of reasons that metadata and metadata standards are important:

- Modernising Government calls for better use of official information, joined-up systems and policies, and services designed around the needs of citizens;
- Considerable work has already been done to standardise government information systems so they can be accessed easily from central portals;
- New systems for the handling of electronic records are being devised. Official records will not always be stored in paper format;
- Metadata makes it easier to manage or find information, be it in the form of web pages, electronic documents, paper files, databases or any other media;
- For metadata to be effective it needs to be structured and consistent across organisations; and
- The e-GIF (e-Government Interoperability Framework) is mandated across all government information systems. By association, so is the e-GMS.

The metadata are of little use, of course, without e-registries in which the metadata can be published and found, but it is also true that e-registries are only as good as the metadata registered in them.

When metadata are published in a registry, semantic interoperability needs to be considered. Computers are entirely devoid of imagination, so automated processes involving text and numbers require that the text and numbers be provided in a carefully structured format. Also, "dead end" and "cul-de-sac" may have identical meanings, but most software cannot accommodate such semantic details. OGC's OpenGIS Specifications for interfaces solve the technical interoperability problem: passing processing instructions and results between dissimilar software systems. But semantic interoperability demands that people cooperate in data coordination activities. Data coordination involves:

- Harmonisation of data models (how the data is structured in terms of the words and numbers used to describe geographic features); and
- Harmonisation of metadata schemas (the data model, plus details that include information about when and how the data was collected, what words mean, what updates have been made)

The Web provides a wonderful tool for managing geospatial semantics issues: the eXtensible Markup Language (XML). XML is a flexible and powerful means of encoding data in text for programmatic manipulation. Although it started out as a language for "marking up" or encoding a document for presentation on a Web page, it has quickly evolved into a mechanism for general data description. As plain text, XML can be read and understood by data managers and can equally be parsed by software programs. XML is easily transformed by simple programs and it is relatively easy to integrate and combine XML-based data from many disparate sources. Most ordinary Web browsers include software that can "parse", or process, text that is structured in XML.

Today, XML is used in a variety of industries including finance, chemistry, e-business, document publishing, multimedia, telecommunications, graphics, and e-government. Each of these domains is developing and reaching consensus on an XML *namespace,* which is a carefully defined set of terms, a *vocabulary* or *ontology,* based on the communication needs of the domain. Like these industries, individual organizations and information-sharing communities are developing their own XML namespaces.

Many of these XML namespaces are being developed in a coordinated international process as "ebXML [electronic business using XML] specifications" (see http://www.ebxml.org/). Development of ebXML specifications is an on-going effort sponsored by OASIS and UN/CEFACT. ebXML enables enterprises of any size, in any location to meet and conduct business through the exchange of XML-based messages. OGC's work on catalogues and registries takes into account current work being done by OASIS and UN/CEFACT.

OGC's Geography Markup Language (GML) specifies an XML namespace suitable for very many geospatial data requirements. In particular, its standard way of expressing locational geometry is applicable in virtually all situations. However, GML and XML do not by themselves solve the problem of two data producers whose data schemas (and metadata schemas) present the same kinds of information in different sequences and structures and whose data dictionaries provide different type names for the same type of geographic feature.

With respect to data coordination, it is useful to think about two classes of data:

- **Base data**: There are several generally useful data layers that ought to have a single data model and metadata schema across Europe (and ultimately, across the world). These include layers such as elevation, water bodies, political boundaries, transportation, surface geology, vegetation and land use. A standard data schema and metadata schema (which includes the data schema) ought to be agreed upon by appropriate authorities, and then all such data can be converted to the standard schemas.
- **Information Community data**: "Information Communities" are groups of people whose profession, discipline, industry, region, nation, government mission or other common interest, including their natural language causes them to have special shared requirements for naming geographic features and for representing relationships among these features. Their data models and metadata schemas ought to conform to base data models and schemas as much as possible and, beyond that, committees within these communities ought to work toward harmonisation within their communities, with consideration for the needs of data sharing partners outside the community.

People will sometimes create profiles of data models and metadata schemas. A profile is a version of a particular schema, adapted from a more normalized schema to meet specific domain needs. Even with such changes, a profile should contain enough standard structure to be publishable on a registry, though the data will be limited in its usefulness to people not working in that domain. A local authority, for example, may have its own profile of a street database.

The power of a Web-based ESDI built on open, standard specifications provides many new incentives for data coordination. When more people can use a data collection, the data collection has more value. When you can use others' data, you do not need to expend precious resources to create your own. When data can be maintained in a single place for use by all, no one needs to store a “stale” copy of the data. In time of emergency, data that has been developed and maintained for routine work can be immediately accessed to save lives and property. And, when semantic translation technologies have matured (in the next 2-5 years) Information Communities will find value in each other’s data.

3. GeoPortals

A **GeoPortal** is an assembly of components that provides a community-wide Web-based access point to resources such as distributed spatial data sources, online geoprocessing services, news, tutorials and tools for collaboration. An ESDI geospatial portal would employ standard software interfaces to connect people through a registry to raster map, imagery and vector feature services set up by providers. It would likely provide specialised glossary, thesaurus, schema examples, XML tools and data committee email lists to support the creation, maintenance and harmonization of schemas. A portal often serves a specific community but it may use a generic user interface that other communities can adapt. A portal usually offers personalised or customised views, serving, for example, a particular Information Community. The workshop participants agreed generally that a European GeoPortal must be based on clear user needs, be multi-lingual to act as a European entry point to available services, and provide links to national portals. To achieve a linking of registries (catalogues) and clearinghouses, existing registries and clearinghouses in different countries need to be extended with OpenGIS Catalogue Services Specification conformant interfaces that enable interoperability.

GeoPortals can:

- Demonstrate what can be achieved by making public sector data more visible and accessible;
- Provide "one-stop shopping" for spatial data and services;
- Provide services that respond to user needs; and
- Identify priority areas for improvements and gaps to be filled.

As a tool to promote SDI development, a GeoPortal provides a measure of progress of SDI development through indicators such as the number of services and catalogues available over time and measures of user feedback.

4. Policy Implications

Political support is absolutely crucial at the national, regional and local levels for the development of these important infrastructures. Developing a coherent policy framework for geographic information is necessary for the whole information society because the SDI is integral of it. Therefore, there needs to be concerted effort applied to implementing the substance of this report and supporting cultural change.

To share data across European information communities will require more than the harmonisation of schemas for semantic interoperability described above. Different nations, regions, agencies, disciplines, professions, industries etc. have different needs and customs with respect to other relevant matters such as data quality, authentication, validation and access security. In the process of implementing new technical capabilities, organisations are often forced to revisit their workflows and policies. Web-based access to geospatial data opens up issues of liability, privacy, freedom of information, intellectual property, public safety and national security. These issues need to be confronted not only by government but also by private sector data producers, data suppliers, software vendors and others.

To open the door to the benefits of a Web-based SDI, and to confront the risks, government agencies, private sector data producers, data suppliers, software vendors and others need to become familiar with the following policy issues for ESDI registries and e-services:

1. E-Services: What kind? What is the business model?;
2. Financing and funding mechanisms;
3. Data quality and interoperability standards;
4. Public sector information, open access, self-governance (liability, privacy, freedom of information, intellectual property rights, e-government, public safety, national security, authentication/security, validation, etc.); and
5. Cross-border situations and multi-lingual aspects.

These are described below.

4.1 E-Services: What kind? What is the business model?

Different jurisdictions, agencies, disciplines, professions, industries, companies etc. have different business models and thus different needs for spatial data and spatial processing. Since the goal is to provide a useful infrastructure for use by all these groups, one must look for their common requirements.

To build e-service models the following aspects have to be kept in mind and should be integrated:

- The different e-service requirements of different stakeholders/users at the European level;
- The provision, price and licensing of services; and
- The funding model for services;

- The market model for services (to develop a free market);
- The liability issues that might apply to services (history and audit).

Different types of e-services need to be considered. One type of service described earlier in this document is the set of registry services that can be used to register metadata, query metadata, and link to data and service servers. These almost certainly would be free services that are part of public registries. Another type would be spatial e-services packaged with specific kinds of data to provide, for example, accessibility information, routing, geocoding or visualisation. Some of these e-services are most appropriately purchased on a per-transaction basis from private sector providers. Others may be seen as generally useful across domains and stakeholder groups and also supportive of private sector value-adding, and thus appropriately provided by government.

A service metadata schema is well documented in the ISO standard 19119. It needs to be reviewed to determine if it is sufficient for spatial e-Services. Designers of the ESDI need to answer the following questions about a service.

- Who is providing the service?
- Who is the service for (target)?
- How is it resourced (finance etc.)?
- Which type/level of service is it?
- What is the service for and what is it trying to achieve?
- Who will regulate the service?
- How does it work, how is it built and maintained?
- How will sustainability be maintained?
- Who are the partners and who do you work with?
- Is the data accessible and guaranteed?

The target groups of e-services are the citizens of Europe, government, business/commerce and research/education.

E-services for e-government can open opportunities for citizens, but those proposing e-government e-services must consider the implications of removing barriers to the awareness and use of geodata and geoprocessing services, as described in the next section.

The question “e-services: what kind?” has not been sufficiently answered. It still has to be clarified how registries could be funded and maintained. The private sector may build registries, but only if they see a profitable business model. If the data is public sector data, then the public sector must put in place the framework necessary (e.g. finance and access and reuse conditions) that will encourage the private sector to provide the required services.

4.2 Financing and funding mechanisms

There is growing awareness about the need for SDIs as resources for good governance and market development. GeoPortals and their underlying registries and e-Services are essential SDI elements, but these cannot be successful without policies and funding that support them. This topic includes both the issue of funding or financing, and the issue of pricing, such as who pays, what is the price of the data, and what are the access policies? The balance between funding via general taxation or user fees is different in different traditions and cultures. Different traditions and cultures also vary in the type and quality of services provided, and this will certainly emerge in policy discussions.

There are four different models for SDI funding.

1. **Government Funding** (funds derived from taxation). Government funding is based on the idea that spatial data produced by the government should be thought of as public infrastructure that ought to be available to everyone "for the good of the commonwealth," e.g., the US Geological Survey.
2. **Private Sector Funding** (funds derived from user fees). Data and services can be sold through contracts or on a transaction basis, to recover cost of production and to distribute that cost fairly among users. This is the "cost recovery" or "data capitalization" alternative. The user contributes to the costs of areas such as collection, updating and quality control, that are not fully funded from elsewhere.
3. **Public Sector Funding** (funds derived from fees charged to public agencies).
4. **Indirect Funding** (funds derived from advertising, sponsorship etc.).

From these basic models, mixed models combining different funding sources can be defined. We are likely to see a combination of public and private funding with significant private input into the policy-making process. Around the world, a number of spatial data collaborative initiatives are emerging. For example, in the U.N., GRID/Arendal is supporting a Spatial Data Consortium within the Consultative Group on International Agricultural Research. In Canada, there are the Geospatial Data Sharing Alliance and the Conservation Spatial Data Consortium. In the U.S., a number of states have spatial data clearinghouses that operate as part of the US Federal Geographic Data Committee's Clearinghouse Activity. Sometimes collaborative initiatives involve creative data sharing agreements. For example, sometimes data, rather than money, is used to pay for other data. Or, base data might be provided at no cost but full market pricing principles could be applied to value added products derived from the base data.

The funding model for the European SDI could be a combined model which encompasses grants (Government Funding) and cost recovery (Private/Public Funding) in an efficient way. But it is likely that most of the funds will have to be Government Funds, at least in the initial implementation stage.

The costs needed ‘on top’ to develop the national SDIs for the European SDI could be funded by EC or by national co-funding.

A European Spatial Data Committee for specifying and promoting the European SDI could be funded by the EC, including the “focal point” for user services and the technical infrastructure associated with this.

Apart from the national funding, additional national level European SDI activities, such as data collection, adoption and metadata catalogues, could need EC co-funding for some activities. The committee for implementation could be involved in this area's development.

4.3 Interoperability standards and data quality

To share data across information communities will require more than harmonisation of schemas for semantic interoperability, as described above. Different jurisdictions, agencies, disciplines, professions, industries etc. have different needs with respect to other matters, including the quality or currency of data. What are the measures of data quality? What quality information needs to be included in the metadata? What about data heritage (source or, if sources, what operation produced the result?), authentication (of heritage) and validation (of accuracy)?

To share data across information communities will require semantic interoperability at the information and service level. Semantic interoperability at the information level requires data coordination. As mentioned above, semantic interoperability at the service level (e-services that provide semantic translation) is not possible yet, but progress is sufficient to ensure that some capabilities will be delivered in products in the next few years.

Semantic incompatibility is a long-term problem for the whole IT community, not only those involved with GI (see W3C's work on the "Semantic Web"). Solutions will be driven by need, and some solutions developed outside the geospatial domain are likely to be very useful for it. Hence, it would be useful to create a typology of needs in the area of spatial data semantics.

Here we list a few:

- There is a general need for harmonisation of data models and metadata schemas for base data;
- There is a need for basic agreement about how to encode data models and metadata in XML schemas. There is a need to explain the standard ways of defining variables in data models and structuring the contents of data sets, the access conditions, the quality etc.;
- There is a need for data element thesauri that enable simple automated semantic translation;
- There is a need to structure registries' interfaces to enable "Google-like" "spatial search engine" searches;
- There is a need for research in the area of automated translation between schemas;
- It is necessary to move towards a geospatial semantic reference framework that will be compatible with the technical interoperability spatial framework defined in OGC. Additionally, an open framework for spatial semantics should be defined. A starting point could be a White Paper, to define problems in different domains, infrastructures, themes and services, to be delivered in Europe's ACE-GIS project

on web services and semantic interoperability (Adaptable and Composable E-commerce and Geographic Information Services http://www.acegis.net.); and

- There is a need to enter a paragraph in the INSPIRE framework to clarify that the importance of semantics needs to be sufficiently recognised.

4.4 Public sector information rights, traditions and security

SDI stakeholders have different practices, customs and assumptions regarding use of spatial data and processing. Opening up the SDI will bring these differences to the forefront. Also, the technologies' new capabilities will spark new debates.

As those in the geospatial information community are aware, advances in geographic information technology raise issues of privacy, accuracy, liability, freedom of information, access security, public safety, national security and intellectual property. These issues directly affect public sector and private sector data producers, data suppliers, and data users, and they indirectly affect everyone. Through expert studies, review meetings and pilot projects, designers of the ESDI and national and local SDIs will learn about and debate the problems, helping to move towards consensus in critical areas. It will be necessary to look horizontally at pan-European EC issues and also vertically at local/national ones.

In the process of defining requirements to exploit new technologies, organisations are often forced to revisit their workflows and policies. Whether the organisations are public or private sector, they are likely to discover that both the benefits and dislocations ripple through their organisations over time.

There needs to be continued discussion of the role of government versus self-management, or stakeholder cooperation, in managing some of these issues. Self-governance of the SDI will in many cases be more effective than governmental rule, but this will demand a commitment to cross-sector cooperation and partnership. Many issues of concern to the GI market will be issues for the whole Information Society.

The INSPIRE framework legislation should set up national bodies with members responsible for the effective implementation of their nations' SDIs including some standards for secure transaction, authorisation etc. The purpose of this is good governance and market development. A similar structure is needed on a pan-European level.

The national body should also have an independent regulator whose role would be to provide coherence in the frameworks that already exist (data protection, freedom of information, access to environmental data, Aarhus etc.). Further on, the regulator should enforce these regulations in collaboration with existing bodies like the Competition Commissioner or an equivalent body in each community. However, it is not the regulator's role to be a data provider or player. Their role applies to everybody including the government and should be independent.

They should help to develop trust, transparency and fair play and should act for the benefit of citizens. This is of particular importance against the background of the still developing GI market which is largely based on PSI (Public Sector Information). As public sector players increasingly act on the market to recover part of their costs, there is a need to ensure that there is no distortion of the market and no abuse of dominant positions and universal service.

4.5 Cross-border situations and language

E-service implementation strategies must rely on and leverage existing and emerging technologies that provide multi-lingual capabilities. Those implementing these technologies should proceed in coordination with groups working on transactions, cross-border issues, IPR issues, security and e-Government.

The cross-border situation forces all countries to work together on the issues of language and "agreement on common information". The multi-lingual aspect should also be synchronised with other countries, because they are all dealing with the same problems.

The meaning of cross-border is "Information that crosses borders". On both sides of the border there is a buffer area of influence on an activity. It is necessary to ensure availability of consistent services on both sides of the border.

The cross-border situation requires two viewpoints. The first one is the "application viewpoint," in which political boundaries are not relevant. Examples include foot and mouth disease, flooding and forest fires. The second one is "legal", where boundaries are relevant.

SDI designers must deliberate on the key applications of common concern and determine if reference and thematic data is required for these applications. The common cultural view on issues such as accessibility and pricing must be considered.

The establishment of e-services and catalogues for cross-border applications requires agreements on standards, reference data, reference system, application, language used and the coordination of programs. Once in place, projects can benefit from available technologies and society can benefit from the projects. It is recommended that money provided for projects should not be spent for capturing data, because this information should be available from existing sources. Instead the money should be used to optimise the infrastructure.

Setting up a strategy for pan-European e-Government services will start in April 2003 and come out as a mandate from the EC. A workshop will be held to formulate the terms of reference for the strategy including general recommendations. The aim is to clarify what the necessary services for European citizens are.

Creating spatial e-content builds the market and creates jobs but it does not necessarily or immediately satisfy the requirements of all end users. The target should be to stimulate value added business and services in cross-border situations in ways that meet the requirements of as many important stakeholder groups as possible.

To achieve this target it is crucial to establish cross-border projects all over Europe, identifying key areas of interest and looking at existing projects (such as Eurosion, a European initiative for sustainable coastal erosion management: http://www.eurosion.org/). Pilots should be started (such as North Rhine Westphalia, in Germany and the Netherlands), showing that cross-border projects can be successful. It is necessary to concentrate first on the regional level, establishing regional structures that facilitate cross-border initiatives.

To identify, for example, legal, cultural and language barriers the cross-border projects should be bundled on a pan-European level. It is proposed to integrate the original projects and pilots into the INSPIRE initiative. This allows participants to learn from non-spatial experiences and establish a two-way dialogue.

5. Recommendations

These are the recommendations of the 21-23 January 2003 GINIE workshop in Munich on Registries and e-Services, as recorded and then formally documented in this paper by OGCE staff. These have been reviewed, revised and approved by the other participants in GINIE.

5.1 General recommendations

The following things should be kept in mind regarding e-services and registries.

- The plurality of usage models at the European level.
- The data and metadata issues related to services.
- The funding model and market model for services.
- The different types of services that might be made available as e-services.
- Liabilities and risks that might be associated with services.
- New e-services should not be judged as a critical success until they are implemented and institutionalised.

5.2 Specific recommendations

The following discussion is a summary of the recommendations discussed in Section 4, “Policy Implications”.

5.2.1 Recommendations regarding which e-services to implement first

1. Each stakeholder group should answer a set of questions (detailed in Section 4), followed by analysis to discover common requirements.
   a) E-Services: What kind? What is the business model?
   b) What financing and funding mechanisms make sense?
   c) What data quality and interoperability standards make sense?
   d) What rules need to be made regarding public sector information, open access, self-governance (liability, privacy, freedom of information, intellectual property rights, e-government, public safety, national security, authentication/security, etc.)?
   e) What must be done to accommodate cross-border and multi-lingual situations?

It is important to ask these questions because different stakeholder groups, such as jurisdictions, agencies, disciplines, professions, industries, companies, have different needs for network access to spatial data and spatial processing. However, two types of e-services are likely to emerge as key multi-application infrastructure elements:

2. The following ought to be implemented broadly and early in the public sector. These are good candidates for public funding.

   a) The set of e-services necessary to implement registries. Users (and their software programs) use these e-services to register metadata, query metadata, and link to data and service servers.

   b) Services to support basic Web mapping: in particular, "cascading map services" and coordinate transformation services that enable Web browsers to obtain and overlay simple raster maps from multiple spatial data servers.

In addition,

3. It is recommended that prototype ESDI GeoPortals be created to provide a focus for SDI-building activities. The services described above would support some of the functionality offered through these portals.

5.2.2 Recommendations regarding financing and funding

1. There is a need to decide how registries shall be funded and maintained. There are four different models for SDI funding:

   a) Government Funding;

   b) Private Sector Funding (funds derived from user fees);

   c) Public Sector Funding (funds derived from fees charged to public agencies); and

   d) Indirect Funding (funds derived from advertising, sponsorship, etc.).

   Mixed models are likely, and creative solutions being implemented elsewhere should be studied.

2. A European Spatial Data Committee for specifying and promoting the European SDI should be funded by the EC. The private sector may build registries and other e-services, but only if they see a profitable business model. If the data is public sector data, then the public sector must establish certain conditions before the private sector will come in.

3. Public organisations should not be allowed to abuse their position to distort the market. Their goal should be to encourage a varied and competitive market in which vendors of many kinds are motivated to meet users' needs for geospatial information.

5.2.3 Recommendations regarding interoperability standards and data quality

To share data across information communities requires two kinds of interoperability:

Technical interoperability, i.e. the ability of diverse spatial processing systems to communicate in real time via shared interfaces.

1. The first recommendation regarding technical interoperability is to be sure that areas such as registries, e-services, map and e-service servers conform to OGC's OpenGIS Specifications. OGC's recent "OpenGIS Reference Model", based on ISO's Reference Model for Open Distributed Processing (RM-ODP), ought to be read and understood by those concerned with ESDI design.
2. It is also recommended that organisations involved in ESDI activities coordinate European organizations' OGC participation to influence the direction of further specification work so that these global standards evolve to serve Europe's needs as much as possible. Continued participation in ISO TC/211 and Web services standards organizations (W3C, OASIS, IETF etc.) is also recommended.

Semantic interoperability. This refers to standards related to data content (including quality issues), naming of geographic features, and schemas for metadata (data about the data). There is a need to:

3. Coordinate and support existing data coordination groups and to focus their efforts on a) base data layers and b) the creation of XML-encoded metadata that can be used in registries. Data coordination needs to become a more visible and organised activity, involving data committees who are the data authorities in their respective information communities. GeoPortals can support this work.
4. Create a typology of European needs in the area of spatial data semantics. Robust e-services that provide semantic translation are not available yet, but research has progressed sufficiently to ensure that some capabilities will be delivered in products in the next few years. A starting point could be a white paper to define problems in different domains, infrastructures, themes and services, to be delivered in the ACE-GIS project on semantic interoperability.
5. Enter a paragraph in the proposed INSPIRE framework to clarify that the importance of semantics needs to be sufficiently recognised.

Note: A third kind of interoperability is machine data format interoperability, i.e., the ability to transfer data between the different formats created by different vendors' software systems. This non-real-time, batch conversion process is time-consuming, error prone and not compatible with online registries and e-services. It is the old way. Though still used, it is becoming unnecessary because of advances in technical interoperability.

6. Work to introduce Web-based solutions that replace data format conversion operations.

5.2.4 Recommendations regarding rights, tradition and security

Advances in geographic information technology raise issues of privacy, accuracy, liability, freedom of information, access security, public safety, national security and intellectual property.

1. It is recommended to have further expert studies, review meetings and pilot projects to advance the level of debate about issues of privacy, accuracy, liability, freedom of information, access security, public safety, national security and intellectual property. The goal is to move toward consensus and action in the most critical of these areas. Cross-sector cooperation and partnership is important.

2. Each national body (see organisational recommendations below) should have an independent regulator whose role would be to provide coherence in the frameworks that already exist (data protection, freedom of information, access to environmental data, Aarhus etc.).

5.2.5 Recommendations regarding cross-border situations and language

1. Move towards a geospatial semantic reference technical framework which is consistent with the technical framework of OGC. Those developing and implementing semantic technologies should proceed in coordination with groups working on transactions, cross-border issues, IPR issues, security and e-Government. The multi-lingual semantic work should be synchronised across Europe, because all countries are dealing with the same problems in this area.
2. Money provided for projects should not be spent for capturing data, because this information should be available from existing sources. Instead the money should be used to optimise the infrastructure. The policy goal should be to stimulate value added business and services in cross-border situations with regard to subsidiarity.
3. To achieve this target it is crucial to establish cross-border projects all over Europe, identifying key areas of interest and looking at existing projects (e.g. Eurosion; http://www.erosion.org). Pilots should be started (such as North Rhine Westphalia and the Netherlands), showing that cross-border projects can be successful. It is necessary to concentrate first on the regional level, establishing regional structures.

5.2.6 Recommendations regarding organisation

1. The framework legislation of INSPIRE should establish a national body, a Spatial Data Committee, in each nation responsible for implementing a national SDI based on standards. These national bodies would play a role in a similar pan-European structure, the European Spatial Data Committee. Each national body should have an independent regulator whose role is:
   - To provide coherence in the framework. Clear agreements should be made on standards, reference data, reference system, application, language and the coordination of programs.
   - To enforce the regulation in collaboration with existing bodies;
   - To help to develop trust, transparency and fair play. It is recommended that the national organizations work closely together to solve problems in ways that provide maximum benefit to Europe.
   - To ensure that the framework benefits all citizens; and
   - To ensure that there is no distortion of the market and no abuse of dominant position and universal service.
2. The cross-border projects and pilots described in 3 above should be integrated into the INSPIRE initiative.
3. There should be close collaboration between geospatial projects and between geospatial projects and other e-Government initiatives.

Appendix A: Towards Geospatial Semantic Web Services

1. Introduction

The proposed INSPIRE (INfrastructure for SPatial InfoRmation in Europe: http://www.ec-gis.org/inspire), the global "Spatial Web", that INSPIRE would be part of, and the World Wide Web, that the Spatial Web is part of, will grow in stages. This paper is an introduction to the geospatial application of technologies that characterize the next phase of the World Wide Web. The Web's inventor, Tim Berners-Lee, calls the next phase the "Semantic Web". So this paper is an introduction to the "Geospatial Semantic Web".

2. The Semantic Problem

What are semantics? *Semantics* in the context of the Web means the intended meaning of objects, their roles, their behaviour and their inter-relationship with other objects and processes within a system. The plethora of research on semantics is mainly focused on three aspects.

1. How to formally represent these semantics in a machine readable form.
2. How to construct Web based reasoning capabilities that utilize this semantically rich environment to enable intelligent Web services.
3. How to encode this semantics in a way so that it can be shared.

The original design of the Web provides information to users from distributed resources that require humans to do the "fine sorting" and the interpretation. Most information on the Web is designed for human consumption. Currently, most Web resources (data and services) do not "describe their semantics" in a way that software can understand, even when derived from a database with a well-defined structure. The Semantic Web will be achieved through 1) enriching data resources and services with semantics and 2) building a set of services that can process those semantics to automate searching and manipulation of data and services in ways that are not possible now.

2.1 Data Semantics

A *data encoding* specifies a particular way to structure the data. Prime examples of encodings include relational schemas, XML DTDs, UML, and Entity Relationship models. Data encoding consists of a finite set of elements that are relevant to the underlying application. The encoding of those elements refers to its *syntax*. Elements are structured and related to each other using well-defined constructs. The constructs organise those elements into objects, attributes and tables in relational schemas; XML elements and attributes in XML DTDs. Each element is associated with a universe of data instances. An element therefore defines a data type. All this defines the *schematic structure* of the underlying application. Behavior, role, and functional relationship of elements is either defined at the application code or left to the users' intuition. This defines the *semantics* of elements.

Data semantics is a familiar problem in the GI industry. Typically, only limited semantic information is captured to describe about real world phenomena. The full meaning and significance of the phenomena being represented are lost at the time of data capture or are represented in the form of mostly human-readable metadata.

So it often happens that users cognitively "add" this information in the course of exploiting these data in an application, or the applications themselves add semantic richness through hard coding in the application code. This lack of sufficient semantic information in data models can lead to the following problems, among others:

- Web services cannot dynamically search and access published data sources. Instead, they must be hard wired to specific data sources.
- Data might be used to solve problems they are not intended for.
- The meaning of the data might be misinterpreted while using it in an application.
- "Semantic heterogeneity" between databases (i.e., differences in meaning and significance) inhibits data sharing.

Through lack of attention to semantics, data development efforts often yield data that is much poorer in information than it needs to be, which severely limits data users' ability to apply those data to effective, multi-discipline decision-making as envisioned in the proposed INSPIRE and in GINIE. For example, data collected for timber management may not easily be reused for mobility or cover analysis. Worse still, we often undergo laborious data conversion at the level of data schemas. These problems have serious, costly consequences for any enterprise. One important consequence of the widespread lack of semantics is that it is exceedingly difficult to reuse data between agencies within the same member state or across member states within the European Union.

2.2 Service Semantics

Service semantics is a relatively new problem, unfamiliar to most users of geoprocessing software. First, what are Web Services? A formal definition of a Web service may be borrowed from IBM's tutorial on the topic: *Web services are a new breed of Web application. They are self-contained, self-describing, modular applications that can be published, located, and invoked across the Web. Web services perform functions, which can be anything from simple requests to complicated business processes…. Once a Web service is deployed, other applications (and other Web services) can discover and invoke the deployed service.*

For many reasons, Web Services are gaining popularity among architects and developers of information systems. Web Services are the key to realising the full vision of Spatial Data Infrastructures. Thus, OGC's members are developing *OGC Web Services* to bring geoprocessing into this new open distributed computing paradigm.

Currently, you can search for a service on the Web by its name and program its URI (Uniform Resource Identifier) into a Web based program. Existing protocols make it work, once you have set it up manually. But semantic interoperability is not yet in place, so a program cannot seek out a Web Service based on its functionality, role, relationship to other service or behaviour, and then automatically invoke it and deliver the result to you the user. To achieve true semantic dynamic discovery and semantic dynamic binding to a Web Service (by a client, agent or other Web Service), Web Services must provide machine understandable semantics. Current Web Service infrastructure development focuses on building the following technology to allow dynamic service discovery and binding (see Figure 1: The Main Characteristics of Web Services Architecture).

SOAP (Simple Object Access Protocol)

SOAP is a protocol specification that defines a uniform way of passing XML-encoded data. In also defines a way to perform remote procedure calls (RPCs) using HTTP as the underlying communication protocol.

SOAP arose from the realisation that no matter what the current middleware offerings are, they need a Wide Area Network wrapper. Architecturally, sending messages as plain XML has advantages in terms of ensuring interoperability. The middleware players seem willing to put up with the costs of parsing and serialising XML in order to scale their approach to wider networks (see http://www.w3.org/2000/xp/).

Submitted in 2000 to the W3C as a Note by IBM, Microsoft, UserLand, and DevelopMentor, the further development of SOAP is now in the care of the W3C's XML Protocols Working Group. This effectively means that SOAP is frozen and stable until such time as the W3C Working Group delivers a specification.

UDDI (Universal Description, Discovery and Integration Service)

UDDI provides a mechanism for clients to dynamically find other Web services. Using a UDDI interface, businesses can dynamically connect to services provided by external business partners. A UDDI registry is similar to a CORBA trader, or it can be thought of as a DNS service for business applications. A UDDI registry has two kinds of clients: businesses that want to publish a service (and its usage interfaces), and clients who want to obtain services of a certain kind and bind programmatically to them.

WSDL (Web Services Definition Language)

WSDL provides a way for service providers to describe the basic format of Web service requests over different protocols or encodings. WSDL is used to describe *what* a Web service can do, *where* it resides, and *how* to invoke it. While the claim of SOAP/HTTP independence is made in various specifications, WSDL makes the most sense if it assumes SOAP/HTTP/MIME as the remote object invocation mechanism. UDDI registries describe numerous aspects of Web services, including the binding details of the service. WSDL fits into the subset of a UDDI service description.

WSDL defines services as collections of network endpoints or *ports*. In WSDL, the abstract definition of endpoints and messages is separated from their concrete network deployment or data format bindings. This allows the reuse of abstract definitions of messages, which are abstract descriptions of the data being exchanged, and port types, which are abstract collections of operations. The concrete protocol and data format specifications for a particular port type constitute a reusable binding. A port is defined by associating a network address with a reusable binding; a collection of ports define a service.

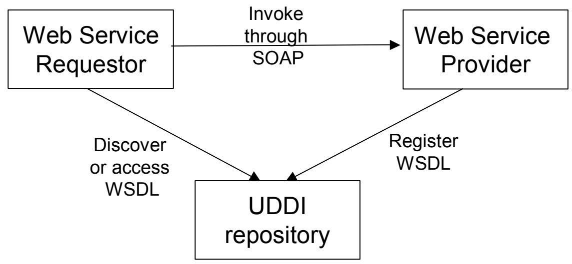

Figure 1: The Main Characteristics of Web Services Architecture

UDDI, WSDL, and SOAP are important steps in the direction of dynamic service discovery and access. However, they only address part of the overall stack of specifications that needs to be available in order to achieve the vision of the Semantic Web. As outlined by the W3C's Semantic Web general architecture (W3C Semantic Web) there are many layers required to achieve automatic Web Service discovery, selection, matching, mediation and composition into complex services.

3. Relationship to the Proposed INSPIRE architecture

At WWW7 (Brisbane, 1997), Tim Berners-Lee outlined his vision of a global reasoning Web. At WWW8 (Toronto, 1998), he articulated the vision of a semantic Web whereby information is given well-defined meaning, better enabling computers and people to work in cooperation. The semantic Web initiative at W3C is gaining momentum and generating technologies and tools that we expect to help bridge the gap between the current standard solutions and the requirement for advanced semantic services.

The Semantic Web requires the use of languages expressing information in a *machine processable* form so that software can provide relevant information or services with minimum human intervention. In this paper we propose the Geospatial Semantic Web, a specialised subset of the Semantic Web. The proposed Geospatial Semantic Web is based on the proposed INSPIRE architecture (see Figure 2: Proposed INSPIRE Architecture).

This describes a system for:

- Exchanging messages;
- Describing Web services; and
- Publishing and discovering Web service descriptions.

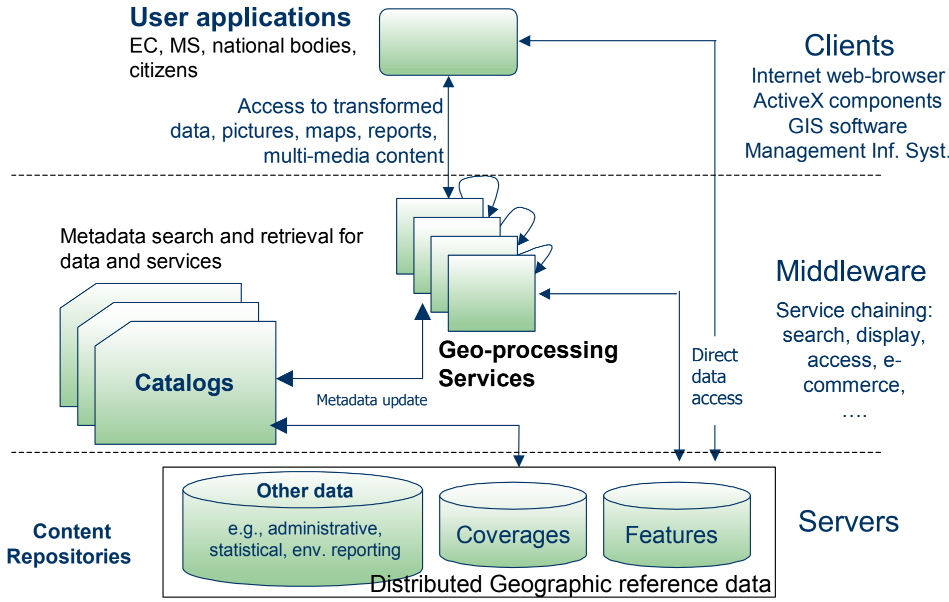

Figure 2: Proposed INSPIRE Architecture

The Web Services architecture defines the interactions between software agents as an exchange of messages between service requesters and service providers. Requesters are software agents that request the execution of a service. Providers are software agents that provide a service. Agents can be both service requesters and providers. Providers are responsible for publishing a description of the service(s) they provide. Requesters must be able to find the description(s) of the services. We recommend to extend the INSPIRE architecture with ontologies and basic semantic services, as explained in sections 4 and 5.

4. Characteristics of the Geospatial Semantic Web

The current Web Services architecture assumes that clients understand the semantics of published services in terms of their input and output parameters. In contrast, the Semantic Web will enable clients to discover and bind to services dynamically without prior knowledge of their syntax and semantics. Semantic Web Services are Web Services identified with a formal description (semantics) that can enable discovery, selection, composition, monitoring, and interoperability. In this case clients can search for services not based on their names and/or syntax, but based on the tasks that need to be executed by the client. Semantic Web Services are found by service requestors that are capable of understanding the description of the service providers. Once the desired Web Services have been found, mechanisms are needed to facilitate combination and sequencing of these services to achieve the tasks at hand. When Web Services are put together, or “chained”, their interfaces need to interoperate. Structural and semantic heterogeneity need to be resolved (see Figure 3: Heterogeneity of Web Services).

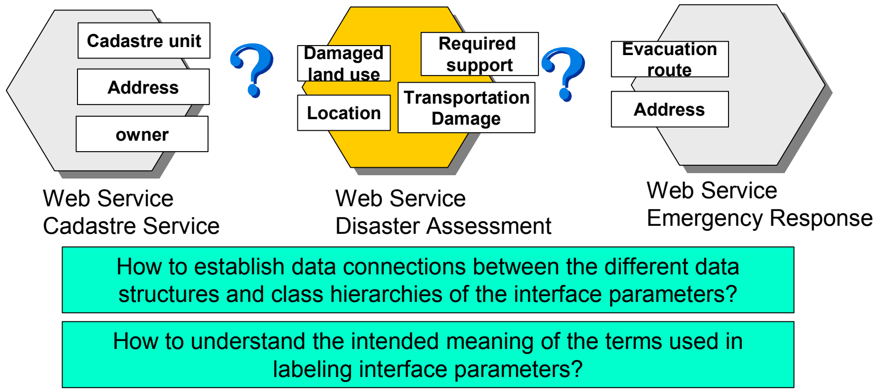

Figure 3: Heterogeneity of Web Services

As shown in Figure 3, structural heterogeneity exists when Web Services use different data structures and class hierarchies to define the parameters of their interfaces. Semantic heterogeneity exists when different service developers use different terms in labelling what are essentially the same interface parameters. The data that is interchanged among Web Services has to have the same meaning on both sides of the interface.

The general idea is that the Geospatial Semantic Web should be aware of its content, understand the interests of its users, and make the best possible use of all encoded information. Open semantic standards must be encouraged within the community to allow the creation of services for knowledge gathering, storage, linking and distribution. Such standards will allow user-friendly client-side utilities to communicate with these services and provide intelligent content selection, composition, processing, and representation functions. Associating meaning with content[1], that is, establishing a layer of machine understandable data about the content, makes interoperable services possible and enables a higher degree of automation and more intelligent applications. The ultimate goal of the Semantic Web is to enable machines to share and exploit data through the use of *ontology*.

Ontology, in the context of the Semantic Web, is the theory of objects and their ties. The unfolding of ontology provides criteria for distinguishing various types of objects (concrete and abstract, existent and non-existent, real and ideal, independent and dependent) and their ties (relations, dependences and predication). Geospatial ontology is the theory of geospatial objects and their ties. Ontological models have the following advantages:

1. They define consensus within an information community on the interpretation of terms;
2. They provide a rigorous foundation for conversion or translation of terms; and

[1] We use the term "content" here to refer to both service and information

3. They provide well-defined terms such as equivalent, inverse, transitive, symmetric, unique property, cardinality, data types.

The following are requirements that the Geospatial Semantic Web community needs to address:

- Automatic Web Service Publishing: once a new resource (service or data) is made available an intelligent software component can harvest the semantic description of the resource and appropriately register it. Semantic description of the Web Service may include a formal description of what the service requires from the user or other software components, what kind of tasks does the service perform, how does the service work in terms of pre- and post-conditions and side effects, and how is the service being used in terms of binding.

- Automatic Web service discovery: A user wants to achieve a task that may require the execution or one of more services. As proposed in several publications about the Semantic Web, a matching service may perform this task[2]. The matching service finds correspondences between the service request and available services.
- Automatic Web Service composition: complex tasks usually require the execution of more than one service. The Geospatial Semantic Web must be able to intelligently analyse task execution requests and compose the appropriate sequence of service requests.
- Automatic Web Service interoperation invocation: once the service(s) are identified, syntactic heterogeneity and semantic heterogeneity need to be resolved.
- Automatic Web Service execution and monitoring: the Geospatial Semantic Web must be able to provide interfaces to monitor the performance of the service. It must provide metrics for the quality of services. One benefit of QoS (Quality of Service) is that it allows client applications to select services based on their quality and also help compose the right set of services.

5. State of the art technology

The Geography Markup Language (GML) will play an important role in the development of Geospatial Web Services. However, it does not provide a full solution to the requirements of the Semantic Web. GML can represent only some semantic properties of geographic objects through its syntactic and schematic structures. Requests for OGC services that return GML need to be aware of these structures via the schema that is defined by a GML application schema. Although one might derive some sort of semantics from the structure of the schema within the context of the application schema, the semantics of each element type is not defined and its interpretation totally relies on

[2] See, for example, "A Semantic Web Approach to Service Description for Matchmaking of Services," David Trastour, Claudio Bartolini and Javier Gonzalez-Castillo, *HP Labs, Filton Road, Bristol, UK*

the implicit knowledge hard coded in application programs. To develop a Geospatial Web with semantics, resources on the Web need to be annotated with structured machine-understandable descriptions of their contents and relationships, using vocabularies and constructs that have been explicitly and formally defined with domain *ontology*. The world view that ontology embodies is a hierarchical description of a set of concepts (is-a hierarchy), a set of properties and their relationships, and a set of inference rules.

In addition to the existing OGC service layers (interfaces defined in OpenGIS® Specifications) and information model layer (GML application schemas), the architecture of the Geospatial Semantic Web must have the following three layers:

1. **The metadata layer**: The data model at this layer contains just the concepts of *resource* and *properties*. The RDF (Resource Description Framework) was developed by the W3C to be the data model for the metadata layer.
2. **The schema layer:** Web ontology languages are introduced at this layer to define a hierarchical description of concepts (is-a hierarchy) and properties. RDFS (RDF Schema), developed by the W3C, is a candidate schema layer language.
3. **The logical layer**: More powerful Web ontology languages are introduced at this layer. These languages provide a richer set of modelling primitives that can be mapped to mostly first order predicate logics. Currently, Web Ontology Language (OWL), developed by the W3C is the most promising logical layer language.

With respect to Web Services, OGC has defined the Web Registry Service (WRS) to describe a standard for an online registry of services and data sources, and for the publishing and dynamic discovery of OGC services offered by service providers. OGC has also defined a Web Service Description Language, which is a derivative from the WSDL. WSDL is an XML language to describe interfaces to Web services registered with a UDDI database, and in the case of OGC it describes Web services registered with WRS.

The Semantic Web will use ontologies to describe various Web resources, hence, knowledge on the Web will be represented in a structured, logical, and semantic way. This will change the way software components (agents) navigate, harvest and utilise information on the Web. On one hand, the Semantic Web is a Web of distributed knowledge bases, where agents can *read* and *reason* about published knowledge with the guidance of ontologies. On the other hand, the Semantic Web is a collection of Web-services described by ontologies. OASIS (Organization for the Advancement of Structured Information Standards) and the United Nations developed ebXML (electronic business using XML: http://www.ebxml.org) to describe business interactions from a workflow perspective. The DAML Services Coalition proposed DARPA Agent Markup Language – Service (DAML-S, DAML-S standards). DAML-S goes beyond ebXML. It facilitates dynamic matchmaking among heterogeneous agents: service provider agents can advertise their capabilities to middle agents; middle agents store these advertisements; a service requester agent can ask a middle agent whether it knows of some provider agents with desired capabilities; and the middle agent matches the request against the stored advertisements and returns the result, a subset of the stored advertisements (Sycara *et al*., 2002).

As can be seen above, we have the technology available for realising the Semantic Web; as we know how to built terminologies and how to use metadata. The whole success of the Geospatial Semantic Web, however, depends on agreeing on common domain standards for each of these layers.

6. Role of European Administrative Bodies and OGCE

It is well recognised within the Semantic Web community that ontologies will play an essential role in the development of the Semantic Web (Barros *et al*., 1998). Various efforts have been devoted to the research of different aspects of ontologies, including ontology representation languages (Corcho and Gomez-Perez, 2000), ontology development (Jones, *et al*., 1998), ontology learning approaches (Maedche *et al*., 2001), and ontology library systems (Ding *et al*., 2001), which manage, adapt, and standardise ontologies.

6.1 Standardisation

OGC and ISO in cooperation with OGCE and various European administrative bodies must lead the standardisation effort to build domain ontologies for data and services. We believe that a basic set of ontological concepts, based on the ISO 19100 series, need to be developed and standardised by ISO and OGC. We call this the Geospatial Backbone Ontology. This ontology will be the basis for more specialised domain ontologies. The Open GIS Consortium must embark on a formal activity to work closely with ISO to build the Backbone Ontology and design a framework for data sharing communities to build their base layer ontologies, (such as, transportation and land use). Currently, a number of ontology representation languages have been proposed (Corcho and Gomez-Perez, 2000) and various ontology library systems have been built (Ding and Fensel, 2001). The question is what would be the standardised ontology for geospatial applications. As mentioned before, we do not intend to advocate a single ontology that is accepted by all involved parties, instead, we advocate multiple overlapping ontologies with mappings between them. OGC, for example, is involved in the Geospatial One Stop initiative in the US to build domain ontologies for transportation themes. Similar efforts can be undertaken in Europe in cooperation with OGCE as well as other data provider coalitions and organisations, including Deutschen Dachverbandes für Geoinformation (DDGI: http://www.ddgi.de/), Digital Geographic Information Working Group (DGIWG: http://metadata.dgiwg.org/) and the National Mapping Agencies.

6.2 Adoption

Since ontologies evolve over time, extending and updating existing ontologies is an important issue. From the technical point of view, this includes searching, editing and the reasoning of ontologies in an ontology library system. From the organisational point of view, European Mapping Agencies and Information Communities play a vital role to promote the use and adaptation of the ontology at all levels across Europe.

6.3 **Management**

The main purpose of ontologies is to enable knowledge sharing and re-use, hence a typical ontology library system supports open storage and organisation, identification and versioning. Open storage and organisation address how ontologies are stored and organised in a library system to facilitate their access and management. Identification associates each ontology with a unique identifier. Versioning is an important feature since ontologies evolve over time and a versioning mechanism can ensure the consistency of different versions of ontologies. We envision a European Geospatial ontology that spans local, regional, national, European, and global levels. These ontologies will be linked, reducing redundancies, and overlaps. European National Mapping Organizations and other European administrative bodies will play an important role in managing these vast ontologies.

6.4 Building geospatial ontologies for Europe

A first step towards the Geospatial Semantic Web for Europe is to lay the foundation to built ontologies for different geospatial domains and formally ground them to existing data sources. Kuhn and Raubal (2003) propose to build a reference ontology for geospatial domains. It is worth mentioning here, however, that it is difficult to construct or standardise a single ontology for a domain that has cross-cultural aspects, as in the EU. Regional Land Use information for member states is a typical example of this. The ontologies vary according to region due to the cultural differences. For example, land use classifications in Italy are different than in Germany. Cross-cultural becomes more pervasive in information space, as research on digital cities shows. On the other hand, it is desirable for users to obtain and publish information based on their own ontologies. Therefore, we favour ontology translation as a mechanism to overcome the semantic problem stemming from cultural differences. We differentiate between this approach and the approach of building unified Ontological models that are acceptable across multiple European cultures. OGCE works on such areas with the EU (see Figure 4: OGCE Semantic Geospatial Activities)

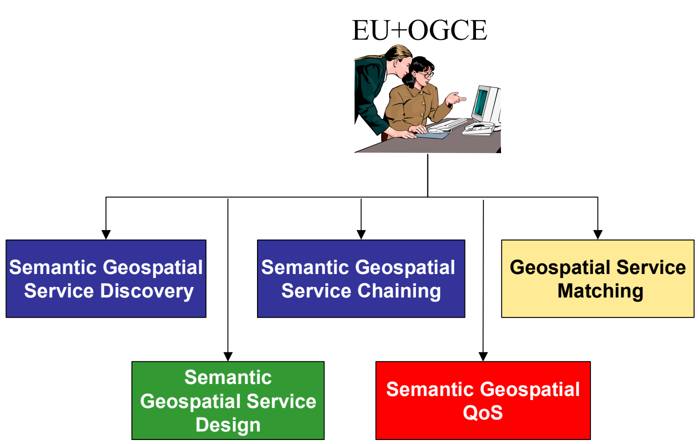

Figure 4: OGCE Semantic Geospatial Activities

Continuing cooperation between the OGCE and the various European administrative bodies will help further the development of the Geospatial Semantic Web Services. We propose the following steps towards the Semantic Web:

1) Build an ontology reference model of the basic geospatial concepts. The Web Ontology Language (OWL) developed by the W3C is a technology that can be used for this purpose;
2) Design a framework to guide the development of domain ontologies and to associate them to data sources. The Web Ontology Language (OWL) can also be used for this purpose;
3) Build ontology for Geospatial Semantic Web services and design a framework to extend this ontology with user-defined domains;
4) Design interfaces to exploit the semantically rich data sources and services developed in the above three steps. This includes publishing, discovery, matching, chaining and binding; and
5) Design and identify the metrics to determine the quality of service of Semantic Geospatial Web Services. And further design the interfaces to implement these.

7. References and Suggested Readings

Hameed, A., Sleeman, D. and Preece. A. (2002) Detecting Mismatches Among Experts' Ontologies Acquired through Knowledge Elicitation. *Knowledge-Based Systems* (special issue) Volume 15 (6) 265–273.

Ankolekar, A., et al., (2001), DAML-S: semantic markup for Web services. in *The First Semantic Web Working Symposium*, pp. 411-430. Heidelberg, Heidelberg: Springer-Verlag.

Barros, F., Goncalves, P. and Santos, T. (1998). Providing Context to Web Searches: the use of ontologies to enhance search engine's accuracy. *Journal of the Brazilian Computer Society* 5(2) 45-55.

Berners-Lee, T. (1998) Semantic Web Road Map. In W3C *IW3C Design Issues*. Cambridge, MA (http://www.w3.org/DesignIssues/Semantic.html)

Brickley, D. and Guha, R. (Eds) (2002), *Resource Description Framework (RDF) Schema Specification*. W3C Working Draft 30 April 2002.

Corcho, O. and Gomez-Perez, A. (2000). A road map to ontology specification languages. in: *Proceedings of the 12th International Conference on Knowledge Acquisition, Modeling and Management, Juan-les-Pins, France*. pp. 80-96. Heidelberg: Springer-Verlag Heidelberg..

DAML-S Standard: http://www.daml.org/services/daml-s/0.7/

Ding, Y. and Fensel, D. (2001) Ontology Library Systems: the key for successful ontology reuse. *The First Semantic Web Working Symposium*, pp. 93-112, Heidelberg: Springer-Verlag Heidelberg.

EbXML standards: http://www.ebxml.org/

Erdmann, M., and Studer, R. (2001). How to structure and access XML documents with ontologies. *Data and Knowledge Engineering*. **36**(3) 317-335.

EU INSPIRE Initiative: http://inspire.jrc.it/

Jones, D.M., Bench-Capon, T.J.M. and Visser, P.R.S. (1998) Methodologies for ontology development, in: J. Cuena (ed.), *Proceedings of the ITi and KNOWS Conference of the 15th IFIP World Computer Congress*, pp. 62-75, London: Chapman and Hall Ltd.

Kuhn, W. and Raubal, M. (2003) *Implementing Semantic Reference Systems* Association of GI Laboratories in Europe 2003. http://musil.uni-muenster.de/index.php?m=publications

Lu, S. (2002). *Semantic Correctness of Transactions and Workflows.* Ph.D. dissertation. Stony Brook, NY: State University of New York at Stony Brook.

Luke, S., Spector, L., Rager, D. and Hendler, J. (1997). Ontology-based Web agents. In *Proceedings of the 1st International Conference on Autonomous Agents*, Marina del Rey, California, pp. 59-68 New York, NY: ACM press.

Maedche, A. & Staab, S. (2001). Ontology Learning for the Semantic Web. *IEEE Intelligent Systems*. **16**(2), 72-79

OGC specifications: http://www.opengis.org/ogcSpecs.htm

OWL: http://www.w3.org/2001/sw/WebOnt/

Pan, J. and Horrocks, I. (2001). Metamodeling architecture of Web ontology languages. in: *Proceedings of the First Semantic Web Working Symposium* (SWWS'01), Stanford University, California. pp. 131-149. Heidelberg: Sprnger-Verlap Heidelberg.

W3C RDF: http://www.w3.org/RDF

W3C Semantic Web: http://www.w3.org/2001/sw/

WSDL Standard: http://www.w3.org/TR/wsdl

SOAP Version 1.2 Part 0: Primerm W3C Candidate Recommendation 19 December 2002. http://www.w3.org/TR/2002/CR-soap12-part0-20021219/

Sycara *et al.*, 2002 [http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/s/Sycara:Katia_P=.html ?]

UDDI Version 3.0 Published Specification, 19 July 2002, http://uddi.org/pubs/uddi-v3.00-published-20020719.htm

Appendix B: Catalogue Mini-Cookbook

1. Introduction

This Mini-Cookbook provides an introduction to implementing catalogues based on the OpenGIS Catalogue Services Specification. In this appendix we refer to these catalogues as "geospatial catalogues". because they describe or refer to (geo)spatial content and/or services. Our goal is to help organizations begin planning how they will put their geospatial data and services online to be discovered and used by others. Like the World Wide Web a decade ago, the value of the emerging "Spatial Web" of networked geospatial data and geospatial processing resources will become increasingly obvious as the number of nodes increases. Just as the Web depends on search engines, the geospatial Web will depend on catalogues, something that can be readily identified in the Architecture Reference Model of the Infrastructure for geospatial Information in Europe (INSPIRE: http://www.ec-gis.org/inspire) (see Figure 1: Catalogues in INSPIRE).

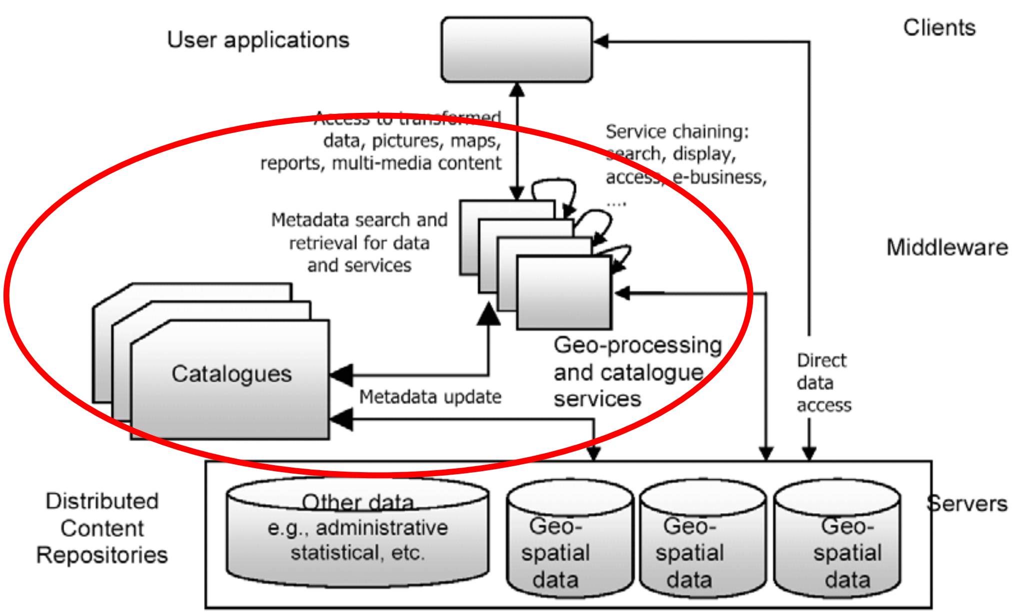

Figure 1: Catalogues in INSPIRE

2. What is a "Geospatial Catalogue"?

Geospatial data and services catalogues allow people and software clients and services to find out what data repositories or services are available and appropriate for their use. The OpenGIS® Catalogue Services Specification (OGC 02-087r3 Version: 1.1.1) explains how Catalogue Services are organised and implemented for the discovery and retrieval of metadata that describes geospatial data and geoprocessing services. The introduction states:

> "Catalogue services support the ability to publish and search collections of descriptive information (metadata) for data, services, and related information objects. Metadata in catalogues represent resource characteristics that can be queried and presented for evaluation and further processing by both humans and software. Catalogue services are required to support the discovery of registered information resources within a collaborating community."

OGC members have reached the consensus that catalogues based on the OpenGIS Catalogue Services Specification should have the following characteristics:

Catalogues enable automated discovery of and automated access to and management of machine-readable metadata describing data that are held in online geospatial repositories (and perhaps off-line data repositories) and also metadata describing online OGC and related Web Services. OGC Web Services are geoprocessing Web Services accessible through interfaces that implement OpenGIS Specifications.[3] Note that a catalogue need not provide all of these capabilities. It may provide only automated discovery of metadata describing geospatial data. The OpenGIS Catalogue Services Specification guides incremental expansion from simple manual geospatial data clearinghouses[4] to catalogues that enable fully automated searches for both data and services. A catalogue supporting highly automated delivery of data and services can also report simply that requested information is available in analogue format or offline media: on paper maps or CD-ROMs, for example[5].

Current OGC Catalogue and OGC Web Services work is shaped by requirements imposed by the IT industry's emerging web infrastructure for “publish, find and bind”. This infrastructure establishes standard ways of:

- Encoding and publishing metadata that describe online services;
- Finding those resources via their metadata; and
- Binding (making the programmatic connection between) service requests and services, including chaining of multiple services from different sources.

The OGC Web Services and related catalogue activities are being designed so that applications and online services can be implemented to automatically find, access and invoke catalogued services. The OpenGIS Catalogue Services Specification v1.1.1 does not fully support access and invocation (binding) of Web Services. The next version (2.0) of the OpenGIS Catalogue Services Specification will migrate to accommodate a Web

---

[3] According to W3C, "A Web service is a software system identified by a URI [RFC 2396], whose public interfaces and bindings are defined and described using XML. Its definition can be discovered by other software systems. These systems may then interact with the Web service in a manner prescribed by its definition, using XML based messages conveyed by Internet protocols."

[4] "A clearinghouse is a collection of institutions providing digital data, which can be searched through a single interface using a common metadata standard." ISO 19115.

[5] Note: The OpenGIS Catalogue Services Specification also supports catalogue implementations that are not web based or Internet based, such as catalogs to resources available in an enterprise via a non-TCP/IP local area network.

Services framework and will use standard approaches as defined by the OASIS ebXML and ebRIM standards (see http://www.ebXML.org). As of May, 2003, the most current document describing OGC's current approach to registries (catalogues) is an OGC-internal Interoperability Program Report: engineering specification, OGC 03-024, entitled, *OWS-1 Registry Service*, which states:

> "The OGC registry information model (ogcRIM) is based on the ebXML registry information model (ebRIM, v 2.1), and thus makes the following assumptions:
>
> 1) all access to registry content is performed through the interfaces defined for the Registry Service;
>
> 2) the information model provides a basis for interoperability by constraining the behavior of the registry (so that it helps the implementor to undestand what to build by defining the kind of metadata being manipulated and how they are interrelated); and
>
> 3) the registry makes use of a repository for storing and retrieving persistent information and shared resources; but no assumptions are made concerning the exact nature of the repository or its location."[6]

The metadata (for both data and services) registered in catalogs must adhere to certain (ISO and OGC) metadata schema standards. Other metadata schema standards and data content standards are not mandatory but may also be useful for certain information community users (see Section 5).

Whichever schemas are employed to structure the metadata, all metadata involved with Web Services are encoded using the eXtensible Markup Language (XML) (see Section 7).

The OpenGIS Catalogue Services Specification defines an SQL-like Common Query Language for search and retrieval of metadata, along with profiles of it for the OLEDB, CORBA, and web computing environments. The web profile uses the ANSI/NISO Z39.50 (also known as ISO 23950) protocol, either on its own Internet port, or via HTTP using XML-encoded requests (see summary in the *Geospatial Interoperability Reference Model*- GIRM).

The content metadata provides information about how to access (view, retrieve and manipulate) the spatial data. This data can be in any raster or vector data format (or even text or video), and the metadata can be held in any data server. However, the data server will not be able to respond automatically to access requests unless the system is online and fitted with interfaces enabling client/server communication. Typically, these will be interoperability interfaces that conform to OpenGIS Specifications.

Spatial catalogues are designed to be distributed. In the geospatial web vision, most catalogues when considered together can be viewed as comprising a single federated network of geospatial catalogues. Owner-imposed access control and security will, quite appropriately, limit people's access to some catalogues and catalogued resources, but access will not be arbitrarily limited by closed, proprietary software interfaces.

---

[6] (Note that this document has not yet received the consensus of OGC's members, that is, it is not an "adopted" OpenGIS Specification.)

The current OpenGIS Catalogue Services Interface Specification describes a catalogue that is *stateful:* servers open a session and "remember" their clients, filling later requests based on earlier ones. However, the Web (linked by the HTTP protocol) is *stateless*: servers treat each request independently. The Web profile of the OGC Catalogue Interface simulates a stateful session using an HTTP "cookie" (see GIRM). Future OpenGIS Catalogue Services Specifications (beginning with v2.0) will adopt more sophisticated methods for managing client-server interaction.

3. Begin with Requirements Analysis and Conceptual Architecture.

The best way to begin in implementing an OGC Catalogue service capability is to develop a very clear and detailed idea about what you (and the end users) want to do, and then to design an architecture and implementation approach based on these requirements.

We recommend that you first review the OpenGIS Reference Model (ORM) (available from OGC's home page, http://www.opengis.org). The structure of the ORM is based on the Reference Model for Open Distributed Processing (RM-ODP, ISO/IEC 10746, see references), a widely used international standard for architecting open and distributed processing systems. The ORM documents the current OGC technology baseline in terms of enterprise, information and technology view points.

The ORM views architecture as a set of components, connections and topologies defined through a series of non-overlapping viewpoints. These viewpoints are defined generally in the RM-ODP and ORM and defined specifically by you in your architecture. The distributed system you are building will have multiple users, developers, operators and reviewers, each viewing the system from their own perspective. The architecture helps to ensure that each view will be consistent with the overall requirements and with the other views. Learn through Pilots and Prototypes.

A key benefit of the ORM is the guidance it provides to build "in an incremental manner". You should first consider the enterprise and information viewpoints in terms of:

- What metadata fields are required for geospatial data to be referenced in the Catalogue;
- What in-house processes or workflows may need to be modified to allow proper capture and maintenance of metadata;
- What queries will need to be supported; and
- What functional requirements for metadata maintenance and capture are required.

Once this work is done, then decide from your requirements which elements of the catalogue system need to be built first and focus on those. Scale the system up over time, focusing on one agency, one program or one technical step at a time.

As in any system integration or software development project, it makes good sense to:

- Involve users in early stage planning as much as possible;

- Record requirements, goals and concepts in detail. It is good to get everything on paper and be sure that stakeholders thoroughly review and discuss the project. UML[7] activity diagrams or use case diagrams can help to formalise this task. (Free tools are available to facilitate the creation of these diagrams.); and

- Let this principle guide you: *Experiment early and often*. In ways that require little investment of time or money, build modest prototypes to learn what works, what does not and why. Disseminate these lessons and use them in early planning. Ask competing vendors to demonstrate interoperability before buying any product. Decide at the outset to use multiple vendors. No vendor providing a complete single-vendor solution will be motivated to provide true interoperability. Get the commitment of the vendors who provided your legacy systems, which they must upgrade with interoperability interfaces if the legacy systems' capabilities are to become network-accessible resources for users of software developed by other vendors.

We recommend that organisations, singly or as consortia, undertake pilot projects to develop local experience with catalogues and other distributed geoprocessing technologies prior to committing to major technology deployments. It especially makes sense for cooperating data sharing communities to work together in pilots, because pilots provide a focused way of addressing the local interoperability issues of greatest concern to those communities.

4. Develop Conformant Data Content Metadata.

Spatial catalogues will not work unless the metadata to be registered in the catalogues conforms to certain standards. Fortunately, data sharing communities have been working together for years to develop metadata standards and data content standards and to produce data that conforms to these standards. Their data coordination efforts will be greatly leveraged by catalogs and OGC Web Services.

Metadata describe data and services so that they may be easily discovered and widely used in an interoperable infrastructure. Metadata are stored in a catalogue and are accessible via catalogue interfaces. The *Geospatial Interoperability Reference Model* (GIRM) provides a good brief overview and the *GSDI Cookbook* (Chapter 4, section on “Relevant Standards”) provides a longer overview of standards for access to metadata through catalogue interfaces, and for metadata content and encoding.

ISO/TC211 (the ISO Technical Committee dedicated to Geographic Information) has published several conceptual models describing geographic metadata, listed below.

ISO 19109 - Rules for application schema provides “... the rules for defining an application schema, including the principles for classification of geographic objects and their relationships to an application schema”. It guides the use of classes,

[7] The Unified Modeling Language™ (UML) is a standard language for specifying, visualising, constructing, and documenting the artifacts of software systems. It simplifies the complex process of software design, making a “blueprint” for construction.

relationships, interfaces, and properties in designing feature schemas for data transfers or transactions.

ISO 19110 - Feature cataloguing methodology defines a standard methodology for an information provider to use in developing a data dictionary, which comprises the bulk of a data content schema. ISO 19110 provides a basis for describing feature types to be pooled across a community of users.

ISO 19115 - Metadata, is a formal schema for geospatial metadata that is intended to apply to all types of geospatial information. ISO 19115 provides a UML model of metadata, based on the US Federal Geographic Data Committee's (FGDC) Content Standard. Its chief purpose is to support profiles, using a small set of required elements and many optional ones. (ISO 19115 is largely harmonised with the Dublin Core and in the US, the FGDC Content *Standard for Digital Geospatial Metadata*.) Many government organisations have developed data and metadata that conform to these standards. All future registered ISO/TC211 metadata profiles must include these core elements, to ensure interoperability and to guarantee productive searches. For the purpose of geospatial catalogues, the schema and core elements of ISO 19115 must be implemented by conforming implementations. A metadata implementation schema (ISO 19139) in progress and will soon offer implementation guidance.[8]

Metadata include, in most cases, the data content schema (or data model), which often (and ideally) is a standard data content schema with a name and a URL-accessible canonical version. A data content schema is a semantic schema, or set of rules and feature type definitions (“data dictionary”), for encoding geographic features' names, attributes and earth-referenced geometry, which are variously referred to as “data content standards”, “data structures”, “data models” and “content models”.

Standard data content schemas enable different people and organisations to maintain consistency among data sets they develop that include the same sets of features, such as the seven sets of features in FGDC's Framework Data (also called “foundation data” or “base data”) layers: elevation and bathymetry, hydrography, geodetic control, cadastral, transportation, governmental units, and digital orthoimagery. However, data content schemas do not apply only to foundation data. Every “application schema” also needs a data content schema. An application schema is a standard data model used by a particular information community[9], such as the community of people who care about flood zones, or the community who care about epidemiology.

In the US, FGDC's Content Standard for Digital Geospatial Metadata defines the content (but not the encoding) of metadata describing geospatial data (XML is the encoding necessary for metadata to be registered in catalogues). This was the starting point for

[8] Those encountering difficulty in obtaining certain ISO/TC211 standards documents can contact the OGC at infor@opengis.org

[9] "Information Communities" are groups of people whose profession, discipline, industry, region, nation, government mission or other common interest including natural language, causes them to have special shared requirements for naming geographic features and for representing relationships among these features.

ISO's Metadata standard (see above). These metadata content standards are used both on their own and as a basis for specialised extensions and profiles. For instance, FGDC has specialised its Metadata Content Standard with Extensions for Remote Sensing Metadata and profiles for Biological Data and Shoreline Data.

Step-by-step tutorials for preparing FGDC metadata are available online from the National States Geographic Information Council (NSGIC) (http://www.lic.wisc.edu/metadata/metaprim.htm) and the Wisconsin Land Information Clearinghouse (WISCLINC) (http://wisclinc.state.wi.us/metadata/metalink.html). A number of tools are available to facilitate metadata development. MetaMaker is a stand-alone metadata development tool for Windows95/NT developed by the U.S. National Biological Service. More information about MetaMaker can be found at http://www.umesc.usgs.gov/metamaker/nbiimker.html. CorpsMet95 is a stand-alone tool for Windows95/NT. It is an upgrade to the original public domain CorpsMet for DOS. It can be downloaded at http://corpsgeo1.usace.army.mil/. Some GIS software vendors also provide metadata development tools, but users of these should check for compatibility with open standards.

5. Set up Service Metadata to Forward Service Requests.

Metadata describing geoprocessing service capabilities rather than geospatial data are increasingly important, because a Web service's public interfaces and bindings are defined and described (in metadata) using XML. As defined by W3C,

> *A Web service is a software system identified by a URI, whose public interfaces and bindings are defined and described using XML. Its definition can be discovered by other software systems. These systems may then interact with the Web service in a manner prescribed by its definition, using XML based messages conveyed by Internet protocols.*

Thus the details of an OGC Web Service resident on Web server can be published in a catalogue, so that a client's (or another service's) request for such a service can lead to the client invoking that service. In some situations, of course, such as an Application Service Provider (ASP) scenario, the client already knows the URI of the service. But, as Web Services roll out, service metadata residing in catalogues will play a key role the new Web Services distributed computing paradigm.

Implementation-level service metadata is an active work area in OGC. At the abstract level, ISO's Services draft (ISO 19119 - Annex C) sketches generic service metadata elements. (See Geoprocessing Services below.) OGC's recent work with Web services has greatly expanded the set of service metadata elements beyond those in ISO 19119. (At the time of this writing, March, 2003, these new services are not yet documented in adopted OpenGIS Specifications. However, much useful information is available in OGC discussions papers (http://www.opengis.org/info/discussion.htm).

6. Use XML and GML.

6.1 XML for Metadata

Metadata schemas structured according to the standards described above establish the structure of metadata documents registered in a catalogue. But the metadata schemas and metadata documents are (usually) encoded in XML. XML (see http://www.w3.org/TR/REC-xml) is a form of structured text that can be read by humans and also processed by software. The metadata can be processed by software, a catalogue parsing thousands of different geospatial metadata documents, each of which describes a different data set, making it is able to report which of those data sets include, for example, data about water bodies in Luxembourg. Also, XML is a key part of the platform for the larger IT industry's Web Services. Metadata describing Web Services (not just OGC Web Services) is encoded in XML.

Metadata are best encoded in XML to take full advantage of the Web Services infrastructure. However, content metadata collections using other data structures, can still support interoperable catalogue searching. By mapping their internal data fields to those of Z39.50's GEO profile, some metadata collections can support FGDC Clearinghouse queries, and indeed, many FGDC Clearinghouse sites still use Z39.50. (http://www.fgdc.gov/clearinghouse/clearinghouse.html) Similarly, an earlier Z39.50 profile, the Catalogue Interoperability Protocol (CIP), supports Committee on Earth Observing Satellites (CEOS) queries across many different metadata collections. CEOS has aligned CIP with GEO where the two schemas overlap. The OpenGIS Catalogue Services Specification was designed to support Z39.50. For structuring XML metadata, an FGDC Metadata DTD (Document Type Definition) is available that conforms to FGDC's Content Standard for Geospatial Metadata.

6.2 GML for Data Model Schemas

Metadata for any kind of geospatial data may be registered in a catalogue. If servers supplying the data are fitted with interfaces implementing OpenGIS Specifications, many real-time operations are possible between diverse vendor's software products. But from the standpoint of maximally leveraging the Web, XML-encoded data model schemas will conform, not only to the relevant ISO standards, but also to OGC's Geography Markup Language (GML 3.0). GML is an XML "namespace," or XML profile, developed, approved and promoted by OGC members as the world's standard XML encoding for geospatial data. GML is an XML namespace, and a particular GML-conformant data model schema is a GML namespace. There are many reasons to have a common XML namespace for geospatial data, including the fact that it provides a common geometry model across all kinds of geospatial data. It is important to note that metadata elements other than the data model schema are not encoded in GML but are encoded in XML.

6.3 GML for Data

A GML data model schema is essentially a "container" for data. Data encoded in GML is structured text that can be processed by software for countless purposes. XML separates content from presentation, so a particular secondary road on a map, or all (or no) secondary roads on that map, can be presented in any colour or line style by the programme displaying the data. Vector and raster geospatial analysis operations can be performed. A brief introduction to GML can be found at:

http://www.Webmapping.org/GMLIntroduction.html.

The current version of the specification can be downloaded from:

http://www.opengis.org/techno/implementation.htm.

7. Learn from Successful Geospatial Catalogues.

The following portals, gateways and clearinghouses all implement the OpenGIS Catalogue Services Specification.

- The **Geoconnections Discovery Portal** (http://geodiscover.cgdi.ca/gdp/index.jsp) is an initiative that is a key access component of the Canadian Geospatial Data Infrastructure (CGDI). It includes a catalogue that enables geospatial resource discovery, evaluation and access. It provides access to maps, satellite images, and other geospatial data provided by Canadian and international organisations. It also provides links to online and offline services of various kinds available or being developed by various organisations. In addition to geospatial data and services, it provides access to professional services and expertise, data publications, papers and presentations, and development resources such as software, hardware, tools, specifications, toolkits, servers, clients, APIs and other portals. It is a very good exemplar for builders of "GeoPortals".
- **Geodata-online** (http://www.geodaten-online.de) is a portal for geodata and geoservices in Germany. It was developed as part of the North Rhine Westphalia (NRW) Pilot Project, an initiative of OGC-Europe and the Landesvermessungsamt (Surveying and Mapping Agency) of the German State of North Rhine Westphalia.
- In New South Wales, Australia, Community Access to Natural Resources Information (CANRI) has developed the NSW **Natural Resources Data Directory** (http://www.canri.nsw.gov.au/nrdd/), which provides a search interface to metadata for natural resources information held within NSW. It is the NSW State node of the Australian geospatial Data Directory. It uses the Z39.50 protocol.
- The UK's Association for Geographic Information (AGI) **giGateway** (http://www.gigateway.org) is the central node for Great Britain's geospatial data. It helps you find and use up-to-date and accurate geographic information from a range of sources and points the user to organisations who supply geographic information products, services and data. It allows visitors to view administrative data from the Office for National Statistics' All Fields Postcode Directory.
- The US **FGDC Geospatial Data Clearinghouse**

  (http://fgdclearhs.er.usgs.gov/FGDCgateway.html) includes over 80 geospatial data servers. The metadata catalogued here include descriptions of the data sets and information on how to acquire or download the information across the Internet through hypertext links in the metadata. It uses the Z39.50 protocol.

- The Netherlands Nationaal Clearinghouse Geo-Informatie **GeoPortal Startup Service** proof-of-concept. (http://www.ncgi.nl/ncgi/).
- South Africa's National Geospatial Information Framework Data Discovery Facility (**SDDF**) (http://www.nsif.org.za/metadata.html).
- From IDEC in Catalunya, Spain: Gestor de Dades Territorials Distribuïdes (**GDTD**) (http://www.geoportal-idec.net/gestor/mapawms/)

8. “Format Standards” and “Transfer Standards”

Historically, different GIS vendors' proprietary data formats have posed a serious interoperability problem for most users of such software. Users persuaded vendors to participate in various efforts to create common data standards, such as FGDC's geospatial Spatial Data Transfer Standard that facilitated bulk conversion from one format to another. These solutions were cumbersome, error-prone and “broke” when the vendors made changes to their internal formats.

OGC has been successful in bringing the industry together to solve the problem in a different way, by developing common interfaces that enable the systems to send and respond to data requests and service requests across networks. The consequences for implementers of geospatial catalogues include:

The metadata document for a data set will tell what file format the data uses. For any geoprocessing system or geospatial database serving geospatial data on the network, however, it really does not matter what file format the data is in, as long as the server is communicating through an interoperability interface that implements the appropriate OpenGIS Specifications.

For data sharing partners who have different software systems, consider prioritising them for upgrades to interoperability interfaces. Perhaps the strategy should be to catalogue metadata for a sample dataset on one server of each software brand. Work to solve the metadata issues and technical interoperability issues on a small scale.

Intermediate data translation (also known as styling) can be handled flexibly using XSLT (XML Stylesheet Transformations); which is especially relevant if data are in the GML format.

Though network-based interoperability reduces the need for data transfer from one format to another, consider the value of converting some datasets to GML, particular where there might be an opportunity to try out the delivery of small feature collections for display on portable devices. GML is easily styled, as mentioned above.

9. Looking to the Future

Future versions of the OpenGIS Catalogue Services Specification will establish a methodology for “stateless” catalogue transactions following true Web Services architecture.

As more people become familiar with geospatial catalogues, XML and GML, data coordination will get a boost because catalogues leverage it and XML-based tools will make it easier. Perhaps on a data coordination portal, publishing an application schema with a feature catalogue for a given data set of common interest, can provide a venue for efficient review and discussion of framework data definitions of use to global, regional, national, and local users. Developed carefully, schemas and feature dictionaries 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.

If different classifications are defined using a consistent set of rules, the ability to map one classification to another and retain the meaning will be greatly increased. This is known as semantic translation. In a few years, current research in semantic translation will result in automated translators that will help make data developed by one information community useful to another.

10. References

Federal Geographic Data Committee (FGDC) (2000) *Content Standard for Digital Geospatial Metadata Workbook* (For use with FGDC-STD-001-1998) Version 2.0, Federal Geographic Data Committee, May 1, 2000.

Nebert, D.D. (ed) (2002) *Developing Geospatial Spatial Data Infrastructures: the SDI cookbook*, Version 1.0, 6 July 2000 and Version 1.1, 15 May 2001. Technical Working Group, GSDI Steering Committee.

Nebert, D.D. (ed) (2002) *Catalogue Services Specification*, (OGC 02-087r3 Version: 1.1.1) (http://www.opengis.org/techno/implementation.htm).

OpenGIS (2003) *Reference Model*. (http://www.opengis.org/info/orm/), OGC, February 14, 2003.

Reed, C. and Nebert, D.D. (2002) *The Importance of Catalogs to the Spatial Web: an OGC White Paper*.

The Geospatial Interoperability Reference Model (GIRM) references geospatial standards and specifications within a highly structured model, to help decision makers choose standards to facilitate interoperable geoprocessing. (http://gai.fgdc.gov/girm/)

Tools for metadata creation: http://geology.usgs.gov/tools/metadata/

Reference Model for Open Distributed Processing (RM-ODP, ISO/IEC 10746). http://isotc.iso.ch/livelink/livelink/fetch/2000/2489/Ittf_Home/PubliclyAvailableStandards.htm

Appendix C: Registries & e-Services Workshop Participants

Name	First Name	Organisation	Country
Bernard	Lars	IFGI Muenster	Germany
Borras	John	Office of the e-Envoy	UK
Buziek	Gerd	CeGI	Germany
Charvat	Karel	Czech Centre for Strategic Studies	Czech Rep.
Craglia	Max	USFD	UK
Corbin	Chris	EUROGI	UK
Donaubauer	Andreas	TU Munich	Germany
Dufourmont	Hans	VlandersGIS	Belgium
Guimet i Perena	Jordi	ICC Catalunya	Spain
Hecht	Louis	OGCE	U.S.A.
Kremer	Claudia	OGCE	Germany
Mueller	Markus	lat-lon	Germany
Ostenson	Olaf	ISO TC 211	Norway
Pichler	Guenther	OGCE	Austria
Rapant	Petr	TU Ostrava, CAGI	Czech Rep.
Rase	Daniel	Eurostat	France
Reem	Dan	Sitescope Limited	UK
Riecken	Jens	Landesvermessungsamt NRW	Germany
Schennach	Gerda	Bundesamt fuer Eich- und Vermessungswesen	Austria
Scholten	Henk	Geodan	Netherlands
Voges	Uwe	Conterra	Germany
Wagner	Roland	Fraunhofer Institute	Germany
Wiegel	Klaus	on-geo	Germany