BENDRASIS PROGRAMAVIMO DOKUMENTAS

EUROPOS SĄJUNGA

NACIONALINĖ ŽEMĖS TARNYBA
PRIE ŽEMĖS ŪKIO MINISTERIJOS

THE BASIS OF AN EU CO-FINANCED PROJECT "QUALIFICATION IMPROVEMENT OF THE LITHUANIAN GEOGRAPHICAL INFORMATION MANAGERS" (NO BPD2004-ESF-2.2.0.-02-05/0143)

Civil servants learning programme

DISTANCE LEARNING OF GEOGRAPHIC INFORMATION INFRASTRUCTURE

Training Material

STRUCTURE OF GEOGRAPHIC INFORMATION INFRASTRUCTURE

GII-08

Vilnius, 2008

MALASPINA
UNIVERSITY-COLLEGE

HNIT - BALTIC

Training material „Structure of Geographic information infrastructure“ (GII-08)

Authors

1, 2, 3, 4 Modules - Brad Maguire and Michael O'Neal Campbell

5 Module- Jurgita Rimkuvienė, Liubov Sverdan and Vaidotas Sankalas (UAB „HNIT-BALTIC“)

Reviewed and edited by

prof. Romualdas Baušys (Vilnius Gediminas Technical University)

Reviewed by

doc. dr. Linas Gabrielaitis (Vilnius Gediminas Technical University)

From English translated and edited by

Astraneta UAB

Study material for civil servants distance learning programme “GII Distance Learning”, which has been approved by a law “RE Approval of Study Programmes for Civil Servants” (No 27V-72; Apr 17, 2007) by the State Service Department under the Ministry of the Interior has been prepared on the basis of an EU co-financed project “Qualification Improvement of the Lithuanian Geographical Information Managers” (No BPD2004-ESF-2.2.0.-02-05/0143)

Beneficiary of the project National Land Service under the Ministry of Agriculture.

Contractor of the project HNIT-BALTIC UAB, Vilnius Gediminas Technical University

Table of Contents

Course Introduction

Welcome to GII-08, Structure of Geographic Information Infrastructure. This course is the eighth out of nine produced for the National Land Service under the Ministry of Agriculture, as part of the Geographic Information Infrastructure Training Program.

Other courses in this program include:

- GII-01: Elements of Geographic Information Systems
- GII-02: Management of Geographic Information (GI)
- GII-03: Standards, Specifications and Metadata for Geographic Information
- GII-04: Applications of a Geographic Information Infrastructure
- GII-05: Geographic DBMS
- GII-06 Geodesy and Cartography for Needs of Geographic Information Infrastructure
- GII-07: Spatial Analysis and Modelling
- GII-09: Web Programming for Geographic Information Infrastructure

In this course, the technical details of how a Geographic Information Infrastructure (GII) are discussed. This course is one of the most technical in the entire program, and builds upon some of the concepts that were developed in previous courses, in particular GII-03, GII-04, and GII-05.

Once you have completed this course, you should have a thorough understanding not only of the overriding concepts behind GII, but you should also have a good understanding of how GIIs work at the level of computer software, hardware, and network infrastructure. This course teaches these topics by examining each in detail, shows how these components work together in a general way, and finally the technical details of the Lithuanian GII are described, as of the time of writing.

The course consists of four major components, including written lecture materials, PowerPoint summaries of the written lecture materials, practical assignments, and a series of exams to test your knowledge. Because students learn in different ways, the material is presented differently in each component. If you find it difficult to understand a particular concept from the lecture notes, it may help to review the PowerPoint slides, or work on the assignment(s) for the module you are studying.

By far, the most detail is provided in the written lecture materials, which include many examples and diagrams to illustrate the concepts presented in this course. The written lecture materials also offer a number of Self-Study Questions to help you think through some of the issues presented in the lecture material, a list of Suggested Readings to help you learn more about the topics covered in the lecture, a comprehensive list of References, and finally a list of Terms Used in the lecture.

Practical assignments have been devised to help you understand some of the details of GII in a more "hands-on" fashion. In spite of the technical nature of this course, the assignments are *less* technical than in other courses in this program because the Lithuanian GII has yet deconstructed and cannot be used for practical assignments. Accordingly, the practical assignments look at some of the individual components of GII, and examine how they have been implemented in other countries.

There will be three exams and one quiz in this course. You will first be given and Inception Exam, which will gauge your understanding of the information presented in this course at the outset. Next, there will be a midterm exam roughly halfway through the course, which will examine the material covered in the first half of the course. Towards the end of the course, a short quiz will be given to help you refresh your memory of some of the materials covered, and finally the Final Exam will cover all of the material in the course, with an emphasis on the material presented after the midterm.

The course is divided into 5 modules. The first module introduces the concepts behind Geographic Information Infrastructures, and looks at how GII is used in different countries around the world. The second module looks at the legal and institutional framework for GII. Module 3 looks at the individual components that make up GII, including Computer Hardware, Computer Networks, and Computer Software. Module 4 examines the integration of these individual components into the "system" that is GII. Lecture 1 discusses the overall design of GII, and Lecture 2 discusses the human and institutional issues surrounding GII development, administration, and maintenance. In Module 5, the design of the Lithuanian GII Framework is presented.

I hope that you find this course challenging, interesting, and rewarding. The concepts that are introduced in this course will hopefully be useful to you on a daily basis as you carry out your activities making E-Government a reality with the Lithuanian Geographic Information Infrastructure.

Hierarchies in GII

Because the GII deals with many different scales of map data, the design of the system requires a thorough understanding of how the different scales interact. There are a number of separate, but interrelated hierarchies that are important to consider in the design of a GII.

Classification Hierarchies are defined to generalise classes of objects as we move to smaller scales, or to specify classes as we move to larger scales. This is the same concept as the idea of having subclasses and superclasses in the object-oriented databases in GIS. Each subclass carries with it all of the attributes of the superclass from which it is created, but because it is more specific, additional descriptive attributes are added. For example, at the scale of 1:2,000,000, we might simply map a body of fresh water as a "lake" and assign it the attributes “pH,” “volume,” and “surface area.” At a scale of 1:1,000,000, we might differentiate these into “natural” or “artificial” lakes. An artificial lake might have a new attribute “creation date” in addition to the generic attributes from the “lake” superclass. The is primarily a "top-down" system, in which classes and subclasses or defined in advance, and the attributes applicable to each class or defined ahead of time, in addition to being inherited from superclass to subclass.

Aggregation Hierarchies are defined to group mapped objects together into increasingly inclusive object descriptions. For example, at a scale of 1:20,000, we might divide moving bodies of water into "creeks," "canals," or "rivers.” At a scale of 1:250,000, all of these might be aggregated into a single "drainage system." In such a case, there is no explicit mapping of attributes from objects at once scale to another, unlike in the system of Classification Hierarchies mentioned above. In this type of bottom-up aggregation, variables tend to be lost unless they are common to all of the types of features that are being aggregated. In a GIS, this might mean that the "length" parameter will be transferred from the individual features to the aggregated feature, but the aggregated feature will be largely graphical in most cases, having very few attributes.

Map Scale Hierarchies apply to map scales defined in national and international map series, as well as to features defined at each particular map scale. Map scale hierarchies are a practical solution to the issues of scale discussed above, and are a response which allows features on large scale maps to be aggregated into features on small scale maps. A number of standard map scales are defined for a nation, which allow topographic features to be mapped at an appropriate scale. Thus, small buildings and houses might be mapped at a scale of 1:20,000 whereas only large buildings would be shown at 1:50,000, the remainder of the buildings would be coalesced into an “urbanized area.”

In general, national map series are defined so that a number of larger scale maps nest perfectly within a single smaller scale map sheet. For example, in Lithuania, each 1:10,000 scale map sheet contains 25 1:2000 map sheets, so each 1:10,000 map sheet is divided into 5 x 5 rectangles (Figure 1). By knowing which rectangle a particular area falls into on a 1:10,000 map sheet, the user can identify a 1:2000 scale map sheet which covers the area at a higher level of detail. Having a nationally complete series of maps that fit into templates in this fashion is one of the prerequisites for a GII (Groot and McLaughlin, 2000).

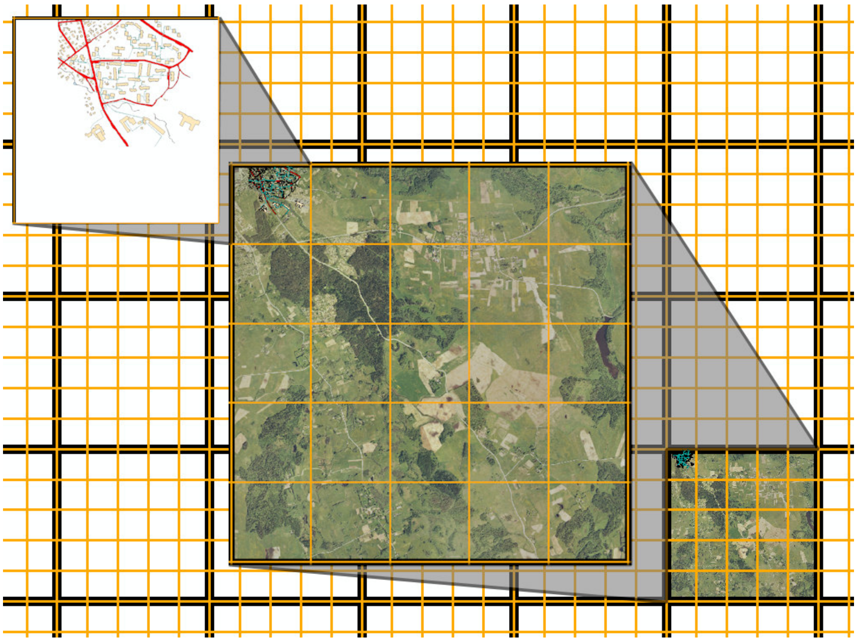

Figure 1. In Lithuania, 25 1:2000 map sheets (top left) fit into each 1:10,000 map sheet (centre), and 25 1:10,000 map sheets fit into each 1:50,000 map sheet (lower right).

One problem with such a system occurs when map projections are changed. Because of the change in the associated datum, the latitude and longitude positions of features change, so map boundaries that are ostensibly defined in terms of latitude and longitude will shift. This means that map sheets defined in LKS-94 will not fit evenly into the 1:1,000,000 grid system used in the UTM Coordinate System of the International Map of the World Series (Figure 2).

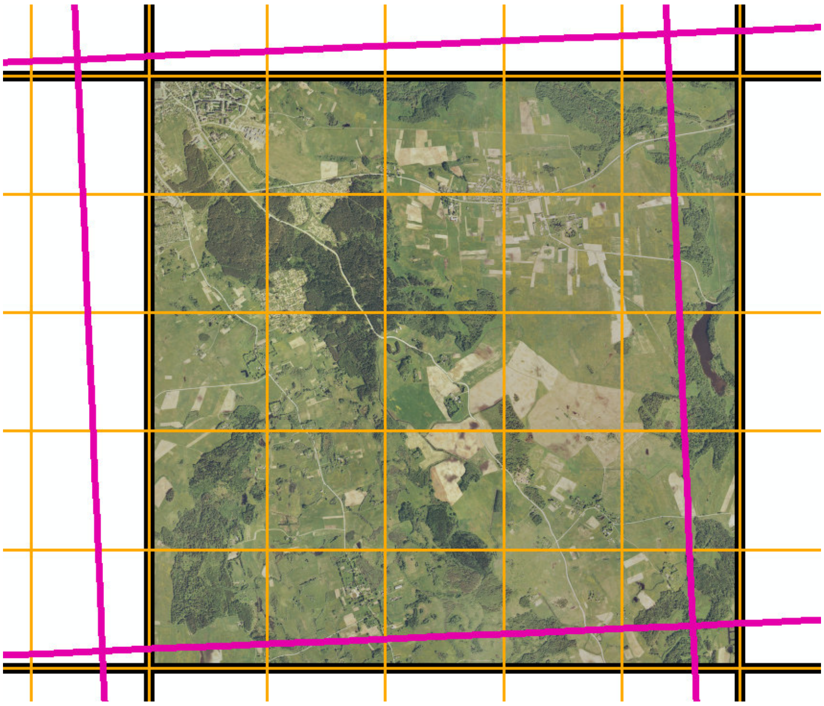

Figure 2. The grids for the UTM coordinate system (purple) are offset and rotated relative to the grids for the LKS-94 coordinate system. For this reason, the Lithuanian grid will not fit perfectly into smaller scale international grids, which are based on UTM.

1 Introduction to Geographic Information Infrastructure

Outline:

1. GIS and GII
2. GII Culture
3. How GII Enables Sharing
4. Building a GII
5. Development of the GII Concept
6. Economic Benefits
7. GII Use Around the World
8. Conclusion

1.1 Geographic Information Systems and Geographic Information Infrastructures

Geographic Information Systems (GIS) were created in response to a number of issues that were facing Cartographers in the 1960's and 1970's. Although paper is a convenient media on which spatial data can be displayed, it suffers from a number of shortcomings that had become evident by the 1970's:

- The original data had to be greatly reduced in volume
- The map had to be drawn extremely accurately, with clear symbolization
- Only a small amount of all available spatial data for an area could be depicted on a map
- It was difficult and expensive to retrieve spatial data from a map
- The printed map was a static, qualitative document (Burrough and McDonnell, 1998)

GIS allowed for better storage and organization of spatial data than had ever been possible using paper maps. It solved all of the above problems because of its approach to handling spatial data. The final report of the 1987 British Government Committee of Enquiry on Handling Geographic Information (also known as the Chorley Report) called GIS "the biggest step forward in the handling of geographic information since the invention of the map" (Masser, 2005).

By making use of computers to handle spatial data, GIS allowed greater volumes to data to be handled more efficiently than ever before. This was a major step forward for public agencies that had to manage spatial information, as Masser points out:

> The advent of GIS and the growing availability of digital map data from the mid-1980s onward made it possible to integrate textual, statistical, and map data within a single system. Such developments dramatically improved the administrative operation of these [public and private sector] agencies and were also reflected in the quality of the services that they provided for the public as a whole. (Masser, 2005, p. 5)

GIS enabled this transformation by logically dividing the process of creating maps into four steps:

- Data acquisition and pre-processing
- Data based management and retrieval
- Spatial measurement and analysis
- Graphic output and visualization (Murai, 2005)

By separating geographic data from the display of that data, GIS made it possible to manipulate the data on its own. Digital data can easily be updated, processed, and transferred between different GIS installations.

The collection and processing of GIS data is a time-consuming and tedious task, so copying and reusing digital data has proven to been very popular. These exchanges of digital data began informally between colleagues, and have gradually increased in importance as increasing amounts of digital data have become available. With this increased importance, the informal nature of the original data exchanges gradually gave way to more formal systems of exchange.

A Geographic Information Infrastructure (GII, also known as a Spatial Data Infrastructure or SDI) is a formal system to share spatial data amongst a broad range of users. In a GII, the collected data are regarded as infrastructure in the same way that a highway might be regarded as infrastructure. Both take large amounts of money to build, and both enable a nation to do things that would not otherwise be possible. In their complete state, they are national assets, but they are assets that must be maintained in order to retain their value. If a highway is not maintained, potholes will develop and bridges will collapse, and if GII is not maintained, the utility of the data that it contains declines or is lost entirely.

1.2 GII Culture

Whereas GIS can be considered to have equal parts hardware, software, methodology, applications, people and spatial data, GII adds to this mix an emphasis on organizations and procedures. Procedures, when enabled by technology, enable organizations and individuals to exchange spatial data. The Global Spatial Data Infrastructure Association website defines GII as:

> A spatial data infrastructure that supports ready access to geographic information. This is achieved through the co-ordinated actions of nations and organizations that promote awareness and implementation of complementary policies, common standards, and effective mechanisms for the development and availability of interoperable digital geographic data and technologies to support decision making at all scales for multiple purposes. These actions encompass the policies, organizational remits, data, technologies, standards, delivery mechanisms, and financial and human resources necessary to ensure that those working at the (national) and regional scale are not impeded in meeting their objectives. (Masser, 2005)

The previous definition touches upon the important role of standardization, which is the single most important factor in a GII. Without standardization of spatial data, there is little point in exchanging anything, because it can be nearly as difficult making use of non-standard data as it is collecting correct data from scratch. A standard for spatial data is often created by a community of users to meet their collective needs. By definition, such a standard contains a well thought-out set of features that meet the needs of all users, and features that are only used for special purposes are omitted.

Martien Molenaar groups data into three classes based on its degree of reusability. Foundation data are general spatial data that have been collected with the idea of reuse, framework data are more specialized, and may be shared within a particular domain of knowledge, for example all environmental disciplines, and application-specific data are so highly specialized that they are very unlikely to be shared at all (Groot and McLaughlin, 2000). A GII concerns itself primarily with the collection and distribution of foundation data, although framework data may be added as a secondary consideration.

For many people, the discipline required to standardize their working procedures does not come naturally, so this is the reason for the emphasis on procedures in GII. Well-conceived procedures help to take the guesswork out of standardizing data, and can actually reduce the amount of time required for data collection, since many of the difficult questions that arise when collecting data, such as how large a parcel of land must be to be collected, have been already answered. It is clear then, that the entire method of operating a GIS is different when the ultimate goal is the reuse of the data than when the ultimate goal is the production of a map. Procedures and standards, quality of work, repeatability of work, and documentation of procedures are paramount in the first case, whereas speed, tailoring data content to what appears best on a map, and cartographic embellishments are emphasized in the second. What we may be seeing is the development of two different *cultures* within GIS – a GII culture and a small-scale GIS culture. Murkund Rao, 2004-05 president of the Global Spatial Data Infrastructure Association, emphasizes this cultural shift:

> Any SDI has to be a national initiative and is envisioned to provide a standardized infrastructure for spatial data holdings of various spatial data generating agencies to be linked into a national system. It is about sharing and providing spatial data and applications services. It is about making spatial data accessible and available, and realizing that the more you make your data available, the more will be the demand for it and also for newer spatial data sets. It is a 'culture' that we will have to imbibe – a culture of standardization and formatting, declaring metadata, sharing, encouraging access and applications." (Masser, p. 235)

It can be difficult to build such a culture, particularly when it is significantly different from the culture that preceded it. Uta When de Montalvo (2003) looked at attitudes towards sharing data in a study of South African GIS practices. The willingness to share was based on three factors:

- Attitudes to data sharing
- Social pressures to share or not share
- Perceived control over data sharing activities (Masser, 2005)

So the move toward GII is not simply an issue of technology, but is an issue of culture as well. Working in an environment that focuses on organizations and procedures, GII specialists will focus on standards, formatting, collaboration, documentation, and above all, sharing of spatial data.

1.3 How GII Enables Sharing

Geographic Information Infrastructures contain technology to make sharing and reuse of data as efficient as possible. Nebert (2004) describes the three most basic functions for a GII. These include:

- Providing a means to discover, visualize and evaluate data
- Hosting of metadata
- Hosting of geographic data and attributes

Using such a system involves first searching for data. Typically, users log in to a GII over the Internet, or sit at a dedicated terminal and search for data by either entering a series of keywords or by locating a study area. Once candidate data has been located, the user can evaluate the quality of the spatial data by viewing it, examining non-graphical attribute data, or by examining metadata, which is a description of the characteristics of a particular data set. Finally, once a suitable data set has been discovered, the data can be downloaded from the GII, in the form of a formatted map or a specialized GIS data format file.

What this leads to is a world in which there is effortless access to free or at cost geospatial data from desktop or mobile computers. How fast we are able to answer spatial questions will be limited only by the speed of the computer and the operator who is making the decision.

This type of environment does not yet exist anywhere in the world. While some new technologies, such as those being developed by the Open GIS Consortium, are helping to make this a reality, they have yet to move out of the experimental stage into the mainstream. Only when users in some country are finally able to make use of geospatial data without even knowing that it is present, will we be able to say that a GII has been completed. Of course, even after this stage, it will be possible to further refine the storage and delivery mechanisms, and increase the number of layers of data stored by the system. As well, the data in the system will always need to be maintained.

This is where GII can take us – Geospatial information that is easily available to all of society. In its most advanced stages, the access to this information is so easy and widespread that the entire population, from children to grandparents make use of this information in some form every day, without knowing it. Just as when you are travelling from place to place, you fail to count the number of bridges that you cross, an advanced GII also becomes “mundane.” GII is infrastructure of a different kind; invisible but ubiquitous, used by everybody.

1.4 Building a GII

Building an infrastructure is not an easy task. Consider again the construction of a new highway. Thousands of workers of all different types are required for such a large undertaking. Planners, Architects, Engineers, Surveyors, Heavy Equipment Operators, Labourers and many other types of workers are all involved in the construction of a new highway. Building a Geographic Information Infrastructure also involves many different people, amongst them Government Leaders, Managers, IT Professionals, GIS Experts, Lawyers, Educators, and Computer Programmers.

Not surprisingly, coordinating so many different kinds of workers is no easy task. If you consider that most of the data in a GII will originate in Government, then it makes sense that the construction of a GII would be a government project.

Ian Masser describes the four key underpinnings of a GII. These are:

- To maximize the use of geographic information. This implies that a GII distributes spatial data broadly, potentially reaching all members of society, and that ultimately, all Foundation and Framework data is made available.
- That governments are the most logical organizations to create GIIs, at the present time.
- The efforts to build GIIs are ultimately user-driven. The primary purpose of a GII is to support decision making, so the decision makers must drive the development of the GII, at least during its early stages.
- As with any infrastructural project, a large range of activities by many different individuals are involved. Such activities include making decisions about data, technologies, standards, delivery mechanisms, institutional issues, organizational responsibilities, national information policies, financial and human resources (Masser, 2005).

We have discussed the types of professionals who are required for the construction of a GII, but we also need to look at who will be involved with the GII once it begins to be operational. The most obvious stakeholder is the Government itself. Since most of the data exchange in GIIs is between government ministries, they are both the main contributors and users of spatial data. Within government, there are primary agencies, such as those involved in topographic mapping and cadastral (property-related) data management, who produce more data than they use, and secondary agencies, such as environmental, geological, or hydrographic agencies, who tend to be net consumers of GII data.

Local government organizations can also make use of data obtained from the GII, to help with municipal mapping, for example. In some cases, local government may also be a contributor of large-scale data to the GII.

Many companies in the commercial sector can make use of GII data. These include information traders and publishers, who take the data from the GII, make improvements to it, and resell the upgraded data as a commercial product. Hardware and software vendors may make use of GII data in their products, or may build their products to make use of data from the GII (a cellular telephone maker adding software to their phones to download maps wirelessly, for example) (Figure 1). Utility and transportation providers are another potential major user of a GII, using it to obtain the location of municipal water and sewer lines prior to building a new natural gas feeder line, for example.

Educational institutions, research institutions, and academics are other potential users of GIIs, downloading spatial data for teaching and research purposes.

Non-governmental and non-profit organizations may also use the data provided by a GII to support environmental and humanitarian projects. In developing nations, development and aid agencies can make use of the data in a GII. In fact, in Ghana, the World Bank has helped to spur the development of GIIs as an aid to national development.

Finally, individual citizens can make use of the data in a GII for everyday activities such as for navigating a vehicle in an unknown area or for personal projects such as producing maps for recreational activities or hobbies. Masser (2005) points out that such individual use of a GII can be used for skill and knowledge development, as well as to help individuals participate in democratic decision-making.

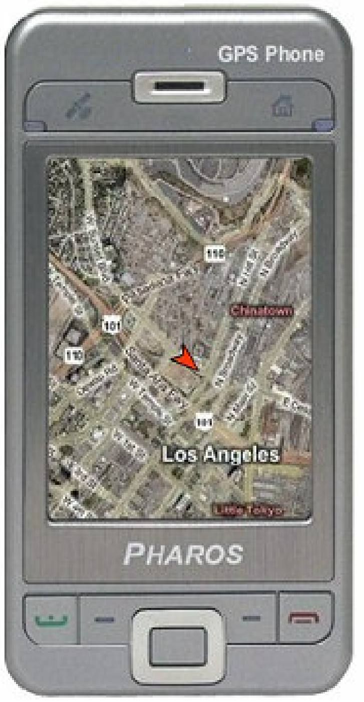

Figure 1. Tools such as this GPS-enabled cellular telephone have the capability to make use of data from a GII

1.5 Development of the GII Concept

The development of Geographic Information Infrastructures can be roughly divided into two phases, one before the advent of the Internet, and one afterwards. It is difficult to overstate the impact of the Internet on society, and it changed some of the fundamental ways that GIIs operated as well. Although the emphasis on standardization and accessibility of spatial data remained unchanged, the advent of the Internet made the delivery of spatial data easier and more flexible than ever before.

The first phase of GII development began about 1980 and lasted until about 2000. During this phase, the focus was on delivering spatial data products. Most of the activities consisted of collecting spatial data and making them available in the GII. Until the Internet became the dominant mode for data transmission, digital data were distributed on magnetic and optical media, such as floppy disks, magnetic tape cartridges, and CDs.

During this phase, GIIs tended to be “monolithic,” since they were characterized by centralized government administration, and single levels of participation, primarily at a ministry-to-ministry level. These early GIIs were explicitly national (or quasi-national in large federal states), and focused on the delivery of data to clients through some sort of information infrastructure.

After about 2000, the availability of the Internet changed the way that data could be distributed, and the way that it could be transferred between individual computers in the GII. Suddenly, it was possible to have a much more flexible network of computers deliver a GII using Internet connections. No longer was there any need for a monolithic centralized GII, but instead a GII could take the form of a flexible network. In fact, it was possible to have more than one GII – you could have smaller GIIs embedded within larger ones. Of course, to the end user, this occurs transparently, so the user never realizes that she is dealing with not only different databases, but entirely different *systems*.

Of course, the advent of the Internet also meant that an individual could access spatial data from any location with an Internet connection, using uncomplicated data access tools. For the governments building GIIs, a large part of the customized infrastructure that was required could be replaced by less expensive computers linked using the Internet.

This new type of GII has coincided with, and encouraged institutional changes in government. Suddenly, it was technically possible for smaller organizations not only to access, but to contribute data to the GII. Assuming that municipalities, educational organizations, non-governmental organizations and individuals are able to meet the rigorous standards for submitting data to a GII, then the importance of GII can increase vastly, with the input of thousands of semi-professionals and skilled amateurs. For an example of how such input can make a positive contribution, visit the web page for Wikimapia, in which individuals can enter information about locations visible in Google Maps (http://www.wikimapia.org/#lat=54.679687 &lon=25.263405&z=13&l=24&m=a&v=2).

The effects of the Internet continue to change GII and government in general. It is clear that many long-established institutional roles are changing because of new capabilities enabled by the diffusion of computer and Internet technology. These changes will affect how GIIs continue to evolve in the future.

1.5.1 Evolution and Adaptation of GIIs

As we saw in the previous section, the concept of GIIs has changed in response to a major technological change, namely the development of the Internet. Geographic Information Infrastructures also change in response to cultural, political, and institutional changes.

In every country where a GII has been implemented, it has been customized to solve local problems. Even once operational, the operation of the GII may change in response to changes in the capacity and role of stakeholders, the addition of new requirements, or improvements in technology.

Political and institutional changes can dramatically affect the structure, or even the existence of a GII. In the United Kingdom, for example, the Association for Geographic Information (AGI) replaced the National Geospatial Data Framework as the lead agency for the British GII project. A 1999 shake-up in the government of the European Community led to a change in direction in efforts to establish a European GII, with the shelving of the GI2000 recommendation to create a working group to coordinate GII efforts in Europe. In its place, new programs were established to support the development of a European GII, such as the Geographic Information Network in Europe (GINIE) and the INfrastructure for SPatial InfoRmation in Europe (INSPIRE).

It is clear that Geographic Information Infrastructures are flexible, complex entities, and that changes in their operational environment can alter their fundamental structure. The GII concept is adaptable to many different environments, and can function at many different levels, from the rudimentary to the highly sophisticated. Because of this flexibility, it is important that the organization managing a GII be cognizant of changes in the technological, political and institutional environments to ensure that the GII structure can be modified to meet changing needs.

1.5.2 Diffusion of GIIs

Since Geographic Information Infrastructures were first proposed, the flexibility of the concept has allowed it to be adapted to fit a number of different scenarios. Originally conceived at a national level, GIIs have been adapted to both larger and smaller scale problems with differing degrees of success. It is clear that GIIs were designed to solve national problems of data distribution and accessibility, allowing national spatial data assets to be managed efficiently. When you have many GIS users and major creators and consumers of geographic data, the GII concept is very logical.

As we move away from this “ideal,” the flexibility of the GII concept has allowed it to be adapted to meet different needs. We can now find some form of Local GII in some states, provinces, and larger cities. Here in Europe, strong integration between national governments has made the creation of a regional (supranational) GII a priority. As this sort of international cooperation becomes more common worldwide, we may even see the creation of a Global GII (Masser, 2005)

Traditional (national) GIIs are created by national governments. There is a strong level of integration imposed by the national boundaries and frequently, the cultural cohesiveness of a people. At a local level, there may be even more cohesiveness than at a national level, but the volume of data to be managed is smaller. At a regional and global level, we tend to have less cooperation, more difficult interactions between groups, and relatively few international data sets, which makes the idea of regional and global GIIs less viable than national ones. At the European level, the medium to long term answer to these issues is to gradually standardise the spatial data products of the nations that make up the EU, so that they can be eventually integrated into a single, seamless, continent-wide spatial database.

At the present time, a number of projects are underway to create regional GIIs. The INSPIRE project in Europe is one such example, as is the Permanent Committee on Geographic Information for Asia and the Pacific (PCGIAP) and the Permanent Committee for the Americas (PC IDEA) (Masser, 2005, p. 182). These projects have only been in existence for a few years, and are many years away from completion. Some international GII initiatives are underway, such as the UN Geographic Information Working Group, the Global Map Program, and the International

Geosphere-Biosphere Programme (Masser, 2005). Even further from completion is a Global Spatial Data Infrastructure (GSDI), which would link all of the regional and international SDIs together into a whole. With the increasing prevalence of international and global environmental issues, and the further development of national, regional, and international GII initiatives, these initiatives are likely to receive increased emphasis in coming years.

The Global Spatial Data Infrastructure (GSDI) is a project that has received a moderate amount of attention over the past ten years. The concept was first developed at a series of conferences beginning in 1996, and the Global Spatial Data Infrastructure Association was founded in 2004. The GSDI concept is still in its infancy, but the GSDI Association is working to develop and promote the concept. Currently, the GSDI Association's focus seems to be on helping to develop national GII associations and fostering international cooperation in the development of GIIs. Possibly its most influential product is the *SDI Cookbook*, which is now in its second edition, and as being updated as new practises are developed.

According to one survey, more than ½ of the World's nations are involved in some form of GII development (Masser, 2005). Unfortunately, because these numbers are based on self-reporting, it is difficult to gauge the accuracy of this figure, particularly for a concept such as GII whose definition has yet to be firmly established because it is so new. Suffice to say that this figure is definitely an overestimation of the number of GIIs that have been developed and are operational.

Another set of numbers helps to show the growth of the GII concept since 1996. In that year, there were 11 GII projects, which increased to 50 by 2000 and 120 by 2003 (Masser, 2005). The following examples describe some of the GII projects in countries around the world.

Portugal, whose GII program was arguably the first in the world, created the National System for Geographic Information (SNIG) in 1990. The Portuguese adopted the 1987 Chorley (British Government Committee of Enquiry on Handling Geographic Information) Report recommendation of having a central organization to spearhead GII research and development, and created the National Centre for Geographic Information (CNIG) in 1990 as well. Possibly as a result of this innovation, Portugal pioneered Internet access to a GII in 1995 (Figure 2). The SNIG Network that was created has helped to modernize local, regional, and national governments in Portugal. The nation has used its GII to promote economic development in its five regions, as well as to promote GIS and economic development in the country as a whole. SNIG is managed by the Ministry of Planning and Territorial Administration (MEPAT), and is funded by the Portuguese government and international agencies (Masser, 2005).

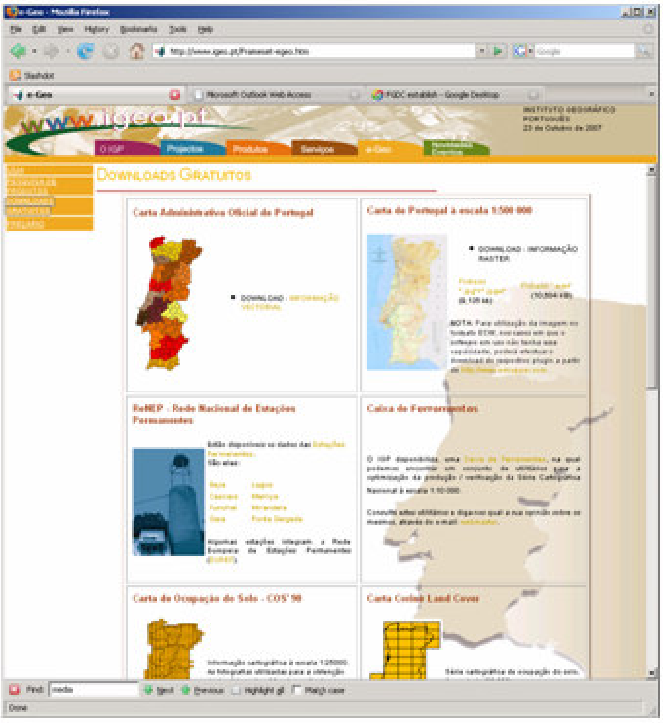

Figure 2. The igeo.pt website contains data from the Portuguese SNIG GII.

In Ghana, government agencies were able to build on work done as a part of the World Bank "Country at a Glance" project that was conducted for Ghana's Environmental Protection Agency (EPA) between 1993 and 1998. This work was adapted and expanded by the EPA to create the National Framework for Geographic Information Management (NAFGIM), which was launched in April, 2000.

NAFGIM is an interdisciplinary and interagency GII that will eventually involve all Ghanaian government agencies. The EPA is the lead agency, and houses the Coordinator as well as the Technical and Administrative staff. A number of working groups are currently involved in developing base cartographic data, biophysical data, socio-economic data, and applications (Figure 3) (Masser, 2005).

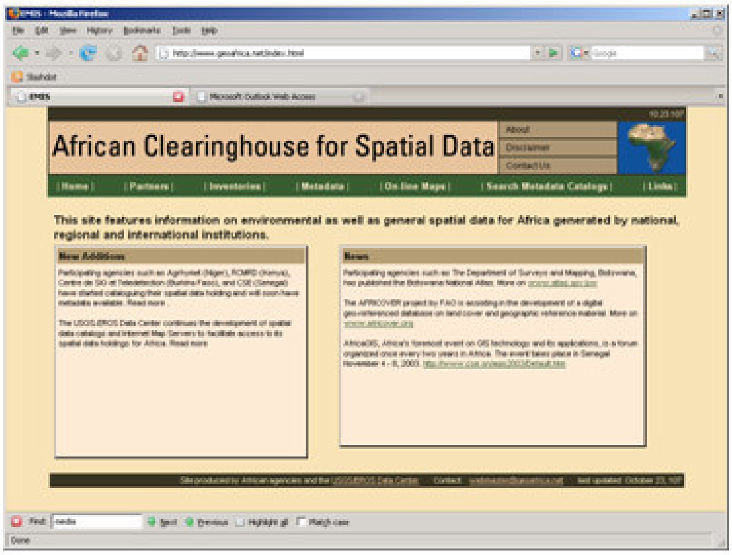

Figure 3. The African Clearinghouse for Spatial Data (http://www.geoafrica.net/) contains GII data from a number of African countries, including Ghana

1.6 Economic Benefits

Geographic Information Infrastructures can have direct or indirect economic benefits. Just as a highway enables intercity commerce, a GII enables an economy to be built around the production, exchange, and improvement, and use of spatial data. Unfortunately, it can be difficult to do the accounting to show the economic benefits of a GII. If a company that delivers supplies to corner stores reduces its delivery time by 5 minutes per day because it uses GII to help route its vehicles more efficiently, can we account for this? What if the same company has 1000 delivery trucks? In such a case, the economic impact is large enough to measure and it can be attributed to the use of GII.

If the benefits are considered in terms of revenues that accrue to the government from the sales of data, then they are likely to be much less than attempts to account for the "gross value added," which looks at the overall economic benefit that results from the distribution of spatial data. Needless to say, estimates of the economic benefits of a GII vary widely, depending on the method of accounting used.

According to the British Ordnance Survey, £79-136 billion was created by GII activities in Britain in 1996; however other statistics show that the government revenues from the sale of spatial data and services were only about $330 million US. Even the second number seems impressive, unless we look at the production costs for the spatial data. Figure 4 shows the balance between revenues and expenditures for the Ordnance Survey between 1979/80 and 1997/98 (Masser, 2005).

In the United States, the National Academy of Public Administration looked at the economic benefits that accrued from the National Spatial Data Infrastructure (NSDI). The estimated that more than $3.5 *trillion* was produced by the 12 core US mapping agencies in 1998. Contrast this with the US Office of Management and Budget estimate that the US NSDI costs (only) $4 billion per year, including staff, overhead, and direct costs. Other estimates of costs have been higher, ranging from $5 billion to as much as $15 billion depending on exactly what is included in the costs for the NSDI (Groot and McLaughlin, 2000).

In Australia, Price Waterhouse took a look at the costs and benefits of the Australian GII. They found a 4:1 cost benefit ratio, and that spatial data products cost the Australian consumer 1/6 of what they would cost without the Australian GII. In total, it is estimated that the system has saved Australians $5 billion over five years. (Masser, 2005).

Admittedly, it is difficult to draw any firm conclusions about costs versus benefits without more concrete numbers, but it appears that, when taken broadly, that the overall economic benefit of a GII may be as much as 1000 times the cost of establishing and maintaining the GII. More pragmatically, it appears that a GII can save a great deal of money, as the Australian numbers show, and has the potential to be self-sustaining economically, as Figure 4 indicates.

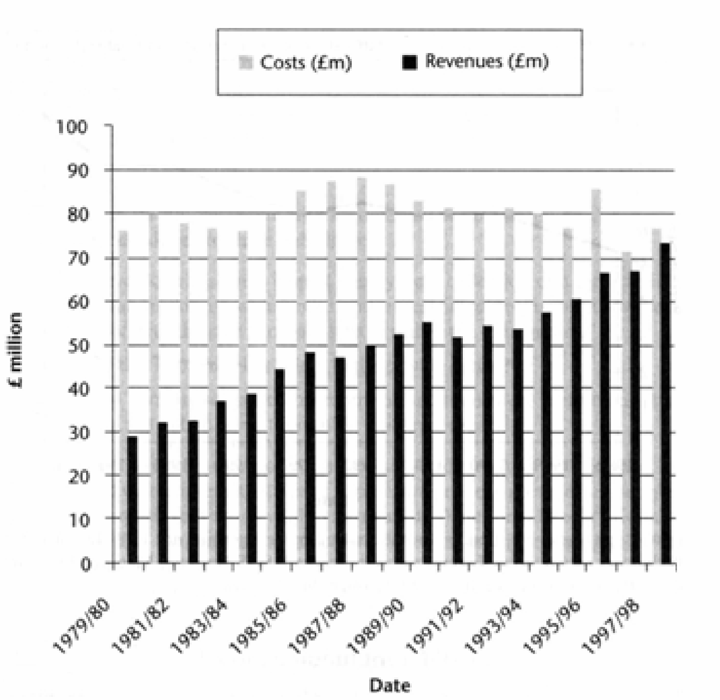

Figure 4. Revenues and Expenditures for the British Ordnance Survey, 1979-1998. Source: Groot and McLaughlin, p. 47

GII Use Around the World

This section examines GII development in three countries: the United Kingdom, Slovenia and Poland. The objective is for the reader to compare the technical political and administrative features of GII in both countries. While GII development in the UK is relatively advanced, the systems in Poland and Slovenia show important improvements towards more sophisticated networking and institutions Important similarities between the there countries will also be observed, as increased governmental interest results in developments across the nation.

1.7.1 GII in the United Kingdom

The UK is composed of three nations, England, Scotland and Wales, and also the province of Northern Ireland. While some infrastructures are national, there are also regional differences. There are slight differences between Scotland and England. The Scottish land registry is different from the HM Land Registry in England and Wales. Decentralization has been promoted, to produce an effective and efficient sharing of information/data sharing among organisations. There is also the adaptation of all government services to electronic systems in 2005, termed e-Government (the UK Office of the e-Envoy, later transformed into the e-Government Unit). These developments increased the recognition of the GI sector as an important actor by the UK government.

The main producers and providers of reference and core thematic spatial data are:

- The Ordnance Survey Great Britain (OSGB), which is the national mapping agency for England, Scotland and Wales, also providing the main reference framework and infrastructure for Great Britain.

- The Intra-Governmental Group on GI (IGGI), which represents central government departments.
- The Improvement and Development Agency for Local Government (http://www.idea.gov.uk)
- The Association for Geographic Information (AGI) (http://www.agi.org.uk), including over 1,000 members from private and public sector organizations, data producers and users. National groups were set up in Scotland (AGIS), Wales and Northern Ireland, after devolution.
- The Geographic Information Panel, started in July 2004 by the government and composed of 12 members, each a senior representative of an organization.

The following organizations and offices also have functions which are also relevant to GII in the UK:

- The Cabinet Office, through its e-Government Unit (http://www.cabinetoffice.gov.uk/e-government/) in respect of e-government policies, and through the Office of Public Sector Information (OPSI: http://www.opsi.gov.uk) for regulating the management of Crown copyright. The Office of Public Sector Information (OPSI) was formed in early 2005, incorporating Her Majesty's Stationery Office, reflecting its role in the UK's implementation of the EU Directive on the re-use of public sector information;
- The Office of the Deputy Prime Minister (ODPM), with responsibility for local government, regional development, planning and some other functions on social issues http://www.odpm.gov.uk ;
- In Scotland the Scottish Executive has similar functions http://www.scotland.gov.uk ;
- In Wales the National Assembly also has similar functions http://www.wales.gov.uk ;
- The major data producers Ordnance Survey, OSNI and Royal Mail. The Director General of Ordnance Survey is official adviser to the Government on GI.

Recently, the Ordnance Survey has led the promotion of GIS development. One development was the National Interest Mapping Services Agreement (NIMSA) of October 13, 1998. Over seven years, the Ordnance Survey will provide improved mapping and other services to the UK government. Several provisions include:

- Provision of a contribution to the cost of maintaining the rural component of the National Topographic Database to the same specification as the more commercially viable rural component.
- Providing a contribution to the cost of maintaining the rural component of the ADDRESS and ROADS datasets to the same specification as the more commercially viable urban component.
- Providing core funding for the depiction of administrative boundaries on the National Topographic Database and thereby supporting the whole boundary revision process in Great Britain.
- Providing core funding for the development of the National Metadata Service in Great Britain to facilitate the discovery of geographic data particularly that referenced in the Annexes of the INSPIRE proposal for a Directive.

- Maintaining the capability to supply the geographical information needs of the educational sector, particularly within the National Curriculum;
- Maintaining the capability to respond to requests from Government and the public for information relating to survey, mapping and geographical information.
- Developing common specifications and standards with OSNI and Ordnance Survey Ireland. The web site http://www.osmaps.org/ provides information and guidance on using mapping and data products from the three National Mapping Agencies responsible for mapping Britain and Ireland. Although they share the name Ordnance Survey they are, in fact, three separate organisations each responsible to their own governments. Each has their own portfolio of products and services, each produced to their own specifications;
- The development of the Digital National Framework, a set of standards and technologies to promote “joined up geography.” This joins all the components of an SDI. DNF is supported by an Expert Group, including central and local government bodies and private sector representatives. There is some collaboration with Ordnance Surveys of Northern Ireland and Ireland.
- Collaboration with the UK Hydrographic Office and British Geological Survey, land registries, local authorities, Royal Mail and the Valuation Office Agency
- The Pan Government Agreement gives the central government access to Ordnance Survey spatial data;
- The Mapping Services Agreement for local government gives local authorities access to Ordnance Survey spatial data;

In Northern Ireland, OSNI developed a common national infrastructure (NIGIS) for sharing spatial information among private-and public-sector organizations. It is based on the Ordnance Survey of Northern Ireland’s topographic database. Other spatial data are referenced to this base. Access to topographical data are obtained from the digital sales office of the Department at Ordnance Survey of Northern Ireland. Electronic distribution has speeded up the process, and a metadata service may highlight datasets that are “fit for purpose.”

Ordnance Survey of Northern Ireland reviewed the NIGIS (Northern Ireland Geographic

Information System), and produced a consultation document “A GI strategy for Northern Ireland” to promote a coordinated cross-governmental approach, standardisation of address data (Pointer), collaboration (cross-border) (Ordnance Survey, OSNI, OSI) and harmonisation. The consultation document was entitled “A Geographic Information Strategy for Northern Ireland”, a draft implementation plan was developed and approved by both the Northern Ireland E-Government Project Board and the Minister responsible for GI policy. The Implementation Strategy, called MOSAIC, defines the collection, storage, management, maintenance, dissemination, analysis and use of spatial information. This includes the creation of an overarching Steering Group, a GI Support Office, and sectoral and project groups including:

- Culture & Heritage;
- Education and Awareness;
- Emergency Services;
- Environment and agriculture;
- Key Datasets;

- Land and property;
- Statistics;
- Health and Social Improvement
- Transport; and
- Utilities and networks.

An Implementation Project Board has been established to oversee the initial implementation of the Sectoral and Overarching Steering Groups.

The Association for Geographic Information (AGI), has two main roles in the context of the SDI in the UK. Firstly as the industry body for geographic information, it forms collective views across the industry on issues of concern, and represents these views to government and others. Secondly, AGI manages and runs the national geographic metadata service (gigateway) under contract from Ordnance Survey, funded by NIMSA. In the first of these roles, AGI has developed a policy on INSPIRE and engages with the relevant UK Government departments to communicate that policy.

The gigateway metadata service is operated by a team of 4 staff within the AGI team (see

http://www.gigateway.org.uk/). The service has 3 components:

• Data Locator - a search engine for geographical information

• Data Directory – a directory of data provider organisations

• Area Search – a lookup service for the different administrative and statistical geographies covering the country

Some of the main landmarks in the development of gigateway have been: The National Geospatial Data Framework (NGDF) was an initiative launched at the AGI '95 event. The emphasis was on a framework of standards, metadata and services. This resulted in two products: a metadata service, and the UK Standard Geographic Base. A website gateway to the services and data was launched: the askGIraffe Data Locator (launched July 2000) and the askGIraffe Data Integrator (launched September 2000). The NGDF was, however, not perceived as being very successful, and in September 2001, the NGDF transferred its functions to the Association for Geographic Information (AGI) and was operated by seconded staff from Ordnance Survey. Responsibility for the askGIraffe Service was transferred to the AGI in April 2002. The contracts of the Ordnance Survey staff ended at this time.

1.7.2 GII in Slovenia

A national GI-association does not exist in Slovenia, but there has been much progress in the production of an effective GII at government, public and private users. In 2004, the Minister of the Environment defined the regulations for the contents and maintenance of the of the spatial data system, including the availability, connectibility and accessibility of data.

There is one major actor for the development of the Slovenian SDI (the Geo-Information Centre), which is part of the Surveying and Mapping Authority of Slovenia (SMA), a branch of the Ministry of Environment and Spatial Planning (MESP). The Environmental Agency of the Republic of Slovenia (EARS), which also belongs to the MESP, is an agent for the development of the Slovenian SDI, focusing on environmental data and applications, and data about spatial planning, such as land use data.

GI at the local level is the responsibility of the municipalities, 12 regional geodetic administrations and 46 government supervised branch offices. Private companies also play an important role in consultancy, data acquisition and application development. The data acquisition is controlled by the Geodetic Activities Act.

The Ministry of Defense, assisted by the Surveying and Mapping Authority of Slovenia, also carries out some technical and development tasks in topography and cartography.

The Surveying and Mapping Authority of the Republic of Slovenia has as its main task providing high quality spatial data for the Slovenian government. It deals with geodesy, mostly the construction and updating of geodetic reference systems, state boundary definition and internal country division, accreditation of geodetic service providers and relevant international co-operation, and land cadastre, the register of buildings and spatial units (addresses), topographical databases and maps. Legal duties involve the dissemination of data, production and management of data catalogues and the supervision of copyright issues.

A major development is the NSDI, built in a centralized top-down approach, with no definite associated legislation, but coordinated by Survey and Mapping Authority, under the Ministry of Environment and Spatial Planning (MESP). There is also a Centre for Geo-Information, which deals with the NSDI's executive work.

The Spatial Planning Act also states that state and municipalities maintain a spatial data system that monitors the both the spatial planning and management.

- Several NSDI-components are in place or under development:
- Reference datasets at various scale levels;
- A comprehensive, standardised metadata catalogue and access service covering geodatasets from a large number of data producers, including SMA;
- A pricing policy catering for non-commercial and commercial re-use of the data;
- An agreement for sharing data within the public sector;
- An ambitious programme to provide more and more advanced spatial data services in line with the GSDI and INSPIRE guidelines.

There is also the development of projects and/or networks, such as the Slovenian Seismic, Water and Meteorological Network of the Geophysical Survey of Slovenia.

The “GeoInformation Centre” (GIC) of the MESP was developed at the SMA in 1991 with several objectives:

- regulation and coordination of GI policy at a national level, and co-operation with other national and international organizations in GI-related standardization, legislation, policy, and legal and organizational aspects of data exchange and distribution,
- development of user services including users requirements analysis, translation of requirements in terms of information processing, technical advice, linking information users and providers and quality support (preparation of quality manual, quality assurance, and quality audits);
- development of metadata services, remote access to metadata catalogues, and data provision through a distributed data warehouse systems;

- raising the awareness of importance of an IT infrastructure, including human resources management, research and development, provision of tools, training, and data integration.

The Geoinformation Centre became part of the Surveying and Mapping Authority in 2001, with unchanged tasks and responsibilities.

1.7.3 GII in Poland

The GII structure of Poland is well developed, and includes structures at governmental level and also public and private organizations. The infrastructure is based upon a basic geodetic network with a uniform set of coordinates, that covers the whole country, to which all measurements are connected, including a total coverage of topographic maps and a well organised system of geodetic technical instructions/guidelines as the obligatory technical standard, and a system of geodetic and cartographic documentation centres. There is also a system for the transferring of the results of geodetic and cartographic works to geodetic and cartographic documentation centres and a legally guaranteed system for registering changes in land use, equipment, cadastre and register of utilities. Utilities are also coordinated, allowing the advance identification of changes in technical infrastructure used on the premises of different organizations. These structures require the skills of highly trained specialists, which are obtained from a network of professionals.

The main organization involved in GII is the Head Office of Geodesy and Cartography.

Head Office of Geodesy and Cartography (GUGiK), associated with the Ministries of Construction and Environment in Warsaw. Other relevant organizations are:

- The Polish Geodetic and Cartographic Service
- The Association of Polish Surveyors;
- The Association of Polish Cartographers;
- Polish Spatial Information Association.
- Ministerial Task Force for Geo-Information Infrastructure,
- National Association of GI Systems Users (GISPOL).

GUGiK has five main functions:

1. Tasks concerning geographic information systems. These concern geographic data resources, data transfer, and development of models related to the functioning of official geographic information systems. Actions taken include:

- The establishment of a Technical Committee for developing Polish Norms in geographic information;
- Preparation of the committee and secretariat for Technical Committee No. 297, following the enactment of the Polish Standardisation Committee, of May 31, 2001;
- The establishment of equivalents of European norms produced by CEN (Committee European de Normalisation/European Committee for Standardisation;
- The establishment of national database of general geographic information and a database of general geographic information at a scale of 1:250,000;

- The implementation of obligations resulting from the Polish participation in the project of EuroGlobalMap (scale of 1:1,000,000, and also for Moldova, Poland, Slovakia, Ukraine and Hungary);
- The initiation and execution of participation in experiments providing experience for designing geographic information systems, and regulations that govern them.
- The testing of a prototype Topographical Database, as defined in the agreement between the Surveyor General of Poland and the Marshal of the Kujawsko-Pomorskie Province for is being developed;
- Technical Guidelines K-2.8 have been prepared and "Principles for the execution of ortho-photomaps in the scale of 1:10,000" issued.

2. Tasks within the field of topographic cartography. Two official topographic maps covering Poland, at scales of 1:10,000 and 1:50,000, based on the international sheet style and the technical regulations of the Surveyor General of Poland. The Head Office of Geodesy Surveying and Cartography (GUGiK) began in 1999, cooperating with the Marshals of provinces, in publishing the Topographic Map of Poland (1:10,000). This included a digital version, and quality control procedures.

3. Tasks in general cartography and geographical onomastics, including the supervision of the "state register of geographical names" and developing the work of the Committee of Standardisation of Geographical Names Outside Poland.

4. Tasks in land surveying (geodesy, development of national geodetic, gravimetric and magnetic network, the measurement of the Polish levelling network. This aims at introducing European vertical reference system to replace the presently used Kronsztadt 86 and satellite techniques for greater mapping accuracy. The Project of Active Geodetic Network ASG-PL. covers the area of the province of Silesia with a geodetic network. It may also add auxiliary components to be used by interested institutions, e.g. police forces, medical services and transportation companies.

5. Tasks in photogrammetry. The collection of photogrammetric aerial photographs in the scale of 1: 13,000 by four specialist companies, and the development of a Central System for Registering Aerial Photographs SEZEL has been functioning. In 2004, the Head Office for of Geodesy and Cartography started the project GEOPORTAL.GOV.PL as a non-profit activity, with a test version in June 2006 (www.geoportal.gov.pl). This makes ortho-images, Digital Model of Terrain (NMT), Raster Map of Poland (RMP), Vector Maps (Vmap) and the General Geographic Data-Base (BDO). The metadata catalogue is based on metadata developed in line with ISO norms (available at: http://meta.geoportal.gov.pl/).

In terms of the legal framework, the geodetic and cartographic activity in Poland is carried out based on the "Geodetic and Cartographic Act" approved by the Polish government on 17 May, 1989. This Law is supplemented by other lower level legal acts which describe in detail the particular items of this law. Preparatory activities are developing draft regulations related to the INSPIRE implementation (in the Geodetic and Cartographic Law). An Interdepartmental Task Force for Geoinformation Infrastructure (2001-2003) and an NSDI Commission (2004-2006) played or are playing an active role.

1.7.4 Lessons Learned

We have seen how GII three countries have been developed, and it can be seen that there are important similarities and differences. Five main lessons may be learned from the study of these examples.

- Consideration of the political system of a country is very important for any assessment of GII.GII infrastructure is required at local levels, and this requires decentralization. In the UK this is achieved through the division of the country into three nations and one province. In Poland and Slovenia, GII is decentralized to the municipalities and provinces. Therefore, GII development is dependent to great extent on the cooperation of central and local government agencies.
- The perception of the relevance and utility of GII by policy makers is an important factor in the development of GII. In all three examples, government policy makers expanded the GII networks, as relevant applications increased. This illustrates the political nature of GII development, and illustrates the key issue for GII users interested in GII development.
- The development of a GII requires technological, legal, economic and managerial developments, within complex networks. As we have seen in each example, new government acts, technology, institutions and networks were created, this increasing the relevance of GII to many new sectors of the countries.
- All the different functions of GeoInformation Science (GIS, remote sensing, cartography, multi-media) may be relevant to the needs of a country and this is reflected in the development of a GII. In all the examples, GII comprised several divisions with several functions, and this multi-functionality increased with increased applications and GII development.
- GII in Poland and Slovenia, although less developed than the UK, is making progress, and these developments reveal the methods required for GII development. The process of development in these countries appeared to begin with centralized institutions and legislation, these gradually expanding as functions, capabilities and relevance increased. The study of countries at different levels of GII development therefore gives important information about the possibilities for developers in different countries.

1.8 Conclusions

In this module, we have looked at the differences between Geographic Information System (GIS) and Geographic Information Infrastructures (GII). GII can be thought of as a set of resources that allow GIS to be used to its maximum effectiveness. We examined the basic functionality of a GII from the user's perspective, and showed how technologies that are currently under development can make the acquisition of spatial data so easy that it becomes an automatic part of daily life.

In the second part of this module, we saw how GII developed in two phases, the first before the widespread penetration of the Internet, and the second afterwards. We looked at how the Internet made it possible to build a networked GII, which might itself be composed of other smaller GIIs. We then looked at how GIIs have spread from their original niche supporting national mapping programs into local, regional, and (potentially) global venues. Finally, we examined the costs and benefits associated with GIIs and noted that the accounting methods used are difficult to justify, but that it does seem clear that GIIs have a great benefit for society as a whole, and are able to return a profit for the lead agency.

In this module, we have looked at the differences between GII in three countries. GII in the UK may be considered the most advanced, and involves the organization of such structures in four units, namely England, Scotland, Wales and Northern Ireland. GII in Poland and Slovenia were seen to be developing, with increased sophistication and capability. We examined the role of the governments, both central and local, and also the role of private agencies. GII was revealed as an important aspect of a nation's development, and a form of technology with international, national, local, social, economic and technical implications.

In each of the three countries covered, we saw how GII is developing, and changing according to the both local needs and government policy. We also saw that there was a process of governmental recognition of the usefulness of GII, involving the development of infrastructures, laws, technical cooperation and responsibilities which were broadly similar, but also reflected the differences between each of the examined countries. The results indicate that GII may be beneficial for any country, provided it effectively relates to the particular context of that country.

Module Self-Study Questions

1. If you always knew your position, could access spatial data from a GII and process it with very little effort, how might this change a trip to a new city? How would it change searching for a restaurant to eat at or planning a trip with a number of stops at different locations?
2. Do you think that having ubiquitous spatial information will cause people to become dependent on it? Before answering, consider how dependent people have become on calculators. On the other hand, consider how much more capable is a person who can use a scientific calculator well than a person of thirty years ago trying to do the same calculations on a slide rule, using logarithmic tables, or doing the operations by hand.
3. How could you effectively assess the state of each nation's GII, without having to visit each of 120 nations?
4. Can you suggest any methods that effectively account for both the costs and benefits of a GII in a manner that represents them in an unbiased manner?
5. What are the main infrastructures required for the development of an effective GII? Comparing the level of GII in the UK or any Western country with that of GII in Poland or Slovenia, describe structures that you think would improve the technical and social environment of the GII. How important are the non-technical issues, such as public access and participation to an effective GII? Are local organizations as important as the central government?
6. Do you think that international cooperation is important for the development of an effective GII? What do you think are the most important forms of cooperation for the development of a GII in countries with less effective GII?
7. How does an effective GII contribute to the development of a country? Describe how a country may benefit in economic, social, technical, environmental and/or military terms. Suggest ways in which the effectiveness of a GII may be improved, to facilitate increased input to one or more of these applications.

Suggested Readings

- GSDI Association Brochure (http://www.gsdi.org/Association%20Information/brochure/2002GSDIbro.pdf)
- McKee, Lance (2000) "Who Wants a GDI?" Chapter 2 in Groot, Richard & McLaughlin, John (2000). *Geospatial Data Infrastructure: Concepts, Cases, and Good Practise*. Oxford: Oxford University Press.
- Ramasubramanian, Laxmi (1999) "GIS Implementation in Developing Countries: Learning from Organizational Theory and Reflective Practice" Transactions *in GIS* 3(4): 359-380.

- GSDI Association Brochure (http://www.gsdi.org/Association%20Information/brochure/2002GSDIbro.pdf)
- McKee, Lance (2000) "Who Wants a GDI?" Chapter 2 in Groot, Richard & McLaughlin, John (2000). *Geospatial Data Infrastructure: Concepts, Cases, and Good Practise*. Oxford: Oxford University Press.

- Ramasubramanian, Laxmi (1999) "GIS Implementation in Developing Countries: Learning from Organizational Theory and Reflective Practice" Transactions *in GIS* 3(4): 359-380.

Assignments

Assignment 1: Geographic Information Infrastructures Around the World

References

- Burrough, P. and R. McDonnell (1998). Principles of Geographical Information Systems. Oxford: Oxford University Press.
- Chang, Kang-tsung (2006). Introduction to Geographic Information Systems, 3rd Ed. New York: McGraw-Hill
- DeMers, Michael N. (2005) Fundamentals of Geographic Information Systems, 3rd Ed. Hoboken, N.J.: Wiley.
- Groot, Richard & McLaughlin, John (2000). Geospatial Data Infrastructure: Concepts, Cases, and Good Practise. Oxford: Oxford University Press.
- Masser, Ian (2005). GIS Worlds: Creating Spatial Data Infrastructures. Redlands, California, ESRI Press.
- Murai, S. 2005. GIS Fundamental (http://www.profc.udec.cl/~gabriel/tutoriales/giswb/vol1/contents.htm, October 22, 2007)
- Nebert, Douglas D. (ed.) (2004). Developing Spatial Data Infrastructures: The SDI Cookbook, Version 2.0 25 January 2004. Global Spatial Data Infrastructure
- Wikimapia web site (http://wikimapia.org, October 23, 2007)
- Groot, Richard & McLaughlin, John (2000). Geospatial Data Infrastructure: Concepts, Cases, and Good Practise. Oxford: Oxford University Press.
- INSPIRE State of Play Reports: Spatial Data Infrastructures in Europe (http://www.ec-gis.org/inspire/state_of_play.cfm)

Terms Used

- Application-Specific Data
- Chorley Report (British Government Committee of Enquiry on Handling Geographic Information)
- Foundation Data
- Framework Data
- Geographic Information Infrastructure (GII)
- Geographic Information Network in Europe (GINIE)
- Geographic Information System (GIS)
- Global Spatial Data Infrastructure (GSDI)
- Infrastructure for Spatial InfoRmation in Europe (INSPIRE)
- Local GII (Local SDI)
- Metadata
- Regional GII (Regional SDI)
- Spatial Data Infrastructure (SDI)

2 The Institutional Framework for GII.

The institutional framework for GII includes the physical infrastructure (hardware, software, buildings, and power supply), the policies and legal framework (government policies, laws, business practices and cultural norms) and the human resource issues (ergonomics, training datasets, institutional framework, technology, standards and financial resources). There are also social issues, namely the user communities and responsibilities, which may influence the location and functions, and political issues, including sharing of data and information with other organisations, nationally and internationally, and economic issues which involve the development of investments, resources for implementation, and operational costs (hardware, software, telecommunication and network). These components interact continuously, this interaction being a requirement for the development of the GII at the national level, and components at the institutional level. All the issues we have examined in the earlier lectures, including GIS methods and practices in different countries are relevant to the institutional framework of GII. This module will examine the organisational, legal and human resource issues of GII institutions. Organisational issues here refer to the physical factors for the effective implementation of GII in an institution. Legal issues include all the laws and policies that enable the establishment, functioning and maintenance of GII, within both specialist GII institutions and multifunctional institutions. We will refer to the previously examined issues that are relevant to understanding the institutional framework of GII. Illustrative examples will be taken from cases in Western and Eastern Europe.

Outline.

1. Organisations.
2. Legal Framework
3. Human Resources.
4. Conclusions.

2.1 Organisational Issues of GII.

2.1.1 Factors for the installation of GII in an institution.

At the institutional level GII may either be the main function of the institution (for example the Geo-Information Centre of Slovenia), or one of many other functions (for example, the Office of the Deputy Prime Minister in the UK). Whichever model is used, GII in organisations is implemented and developed with specific objectives and common justifications. The following benefits are the main factors justifying the introduction of a GII into an institution:

- Geospatial data are better maintained in a standard format
- Revision and updating are easier
- Geospatial data/information easier to search, analyse and represent
- More value added products can be created
- Geospatial data can be shared and exchanged freely
- The productivity and efficiency of the staff is improved
- Time and money are saved
- Better decisions can be made
- Data acquisition and pre-processing is simplified
- Data based management and retrieval speed is increased
- Spatial measurement and analysis performance is improved
- Graphical output and visualization is made better

These benefits may have wider benefits, such as the following:

- Improved social infrastructure leads to a better society
- Environmental infrastructure....better management
- Urban infrastructure.....better life
- Economic infrastructure.......better business
- Educational infrastructure......better knowledge

2.1.2 Requirements for an effective institutional GII.

For an effective GII in any organisation, several factors must be present, both within and between organisations:

- Open data policy. GIS data and information should be accessible by any user, freely or inexpensively and without restriction. Restrictions must be based on the type of data collected, with increased restrictions for sensitive data.
- Standardization. Standards for data format and structure should be developed to enable transfer and exchange of geospatial data.

- Data/Information sharing. In order to save cost and time for digitisation, data sharing should be promoted. In order to foster operational use of geospatial data, information and experience should be shared among users.
- Networking. Distributed computer systems as well as databases should be linked to each other to a network for better access as well as better service.
- Multi-disciplinary approach. Because GIS is a multi-disciplinary science, scientists, engineers, technicians and administrators of different fields of study should cooperate with each other to achieve the common goals.
- Interoperable systems. GIS should be interwoven with other systems such as CAD, computer graphics, image processing, DEM etc.

2.1.3 Users of GIS and GII.

There are four user groups of framework datasets within an organisation and in the environment of the organisation:

1. Primary users (the collector and major users);
2. Secondary users (incidental users for similar purposes as the primary user);
3. Tertiary users (users that use the dataset for other purposes than the purposes for which the information was collected and the dataset created), and
4. End-users.

2.1.4 The components and installation of hardware and software in an institution.

As we explained in earlier lectures, a GII is based on computer hardware and software, people, regulations and managed networks, including procedures which may support the data capture, processing, analysis, modelling and display of geospatial data. Here we list some of the important hardware and software. These vary according to the size and complexity of the organisation, the complexity of the tasks allotted to the GII, and the availability of technology and training. These are relevant to the issues listed above (2.1.2, 2.1.2, 2.1.3). Specialist GII institutions may require more sophisticated inputs, as may institutions with multiple functions networking with several users. The points below are mentioned, so the reader may think through how such inputs may vary according to the problems mentioned in the last few lectures (issues concerning access to technology, networking and international cooperation). The reader may also consider how such facilities may be decentralised (in association with the following section on human resources) the costs and training required for such developments. In the considered examples (UK, Poland and Slovenia) decentralisation was a key issue in the utility of a GII system.

The hardware system. A hardware system is supported by several hardware components.

1. Central processing unit (CPU). This executes the programs and controls the operation of all components. Usually a personal computer (PC) or a workstation is selected for the required CPU or as a server computer.
2. Memory. The main memory is essential for the operation of the computer because all data and program must be in main memory for fastest access.
3. Auxiliary memory. This is used for large permanent or semi-permanent files with slower access. Hard disks, USB flash drives, magnetic tapes, or optical compact disks (DVD or CD-ROM) are used. At least more than 1 G bytes is required for hard disk in GIS.

4. Peripherals. These include the keyboard, mouse, digitisers, image scanners, digital cameras, digital photogrammetric workstations, colour displays, printers, colour plotters, film recorders etc.

The software system. A software system is composed of programs including operating system, compilers and application programs.

1. Operating System (OS). This controls the operation of the programs as well as all input and output.

For PCs: Microsoft Windows is the dominant OS.

For Workstations: UNIX is the dominant OS.

2. Compilers: convert a program written in a computer language to machine code so that CPU can execute binary operation. Commonly used languages include Visual Basic, Visual C++, and C

Geospatial data. This geospatial data includes digitised maps, aerial photographs, satellite images, statistical tables and associated documents. It may be classified into graphic data (geometric data) and attributes (thematic data).

2.2 Legal Framework of GII institutions.

2.2.1 Classification of legal instruments.

Legal issues in GII are concerned with several key issues, related to the perception, classification, utility, cost, demand, and accessibility of the products of the GII. Geospatial data and information may be considered a public good, under freedom of information laws. GIS may also be considered an information utility. In either case, policies and laws are developed to control, regulate and enable public access to products, and also regulate the production of software. In most countries, laws comprise the following:

- Pre-existing laws covering information access, including copyright and patents.
- Privacy laws, applied to both information access, and the use of remote sensing methods such as webcams, security cameras, aerial photography and satellite imagery.
- Laws governing the establishment of new contacts between central and local government, local public sector bodies and private companies and individuals.
- Laws establishing new bodies that expand the capacity of GII.
- Laws that repeal some of the impediments to the smooth functioning of an expanding GII.

2.2.2 Copyright and privacy laws.

In Europe (EU countries) freedom of expression and information is protected by Article 10 of the European Convention for the Protection of Human Rights. This allows the right to receive information without interference from a public authority. However, this does not include an obligation for public bodies to provide information. National laws also protect copyright and patents. Such laws are applicable to the access to GIS data (including tabulated data, satellite and photographic images, and maps). Legal protection of GI data is important, as this prevents the illegal use of the information, and reduces losses to competitors, by reducing access. Legal protection includes:

- Copyright law. This protects intellectual, usually original achievements, and is less applicable to the factual data of GIS, which may be original mostly in terms of how it is organised or presented.
- Unfair competition law. This prevents dishonest practices, such as bribery to obtain data.
- Contractual law. This applies to actions beyond that agreed by contracting parties.
- Right of extraction (EC Directive 96/9/EC on the protection of databases. This is the most relevant to geospatial data, as this gives the maker of the database rights provided there is a strong interest in “either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilisation of the whole or of a substantial part…” of the database (which is defined as “a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means).

The negative aspect of these restrictions is the impact on the user or customer for the GIS information. The restriction on the free flow of information may affect both public and private users. The issue is to balance the use of the laws with the access of the users.

Privacy laws are relevant to remote sensing of images and videography, which may record activities within people's private lives. GII is concerned with both the collection and analysis of data and barriers to the collection of data is a debatable issue. Restrictive legislation may limit or regulate the use of remote sensing equipment such as webcams, security cameras, helicopter-based videography, aerial photographs, satellite images and radar images. As many institutions collect such data, restrictive legislation may regulate institutional operating procedures and limit user access to derived information. Laws include:

- Limitations on central government organisations such as the air force and police in data gathering, these limitations usually based on national, or in Europe, EU level legislation
- Limitations or prescriptions on local organisations, which regulate the use of webcams, security cameras and other small scale data gatherers
- Regulations on the responsibility of different institutions for different levels of data gathering. For example, police surveillance may require a warrant, and the data gathered may be restricted to law enforcement agencies or other authorised organisations. Data gathered by surveillance by private security agencies may be protected by privacy laws.

2.2.3 Institutional and organisational laws.

Laws governing the establishment of institutions define the powers, operating procedure, functions, limitations, contacts and long term goals of the institution. For GII, the relevant functions are defined, which may occupy part or all, of the mandate of the institution. For example in the last module, we learnt that several institutions in the UK, Poland and Slovenia were involved in GII, and this aspect varied according to the mandate of the institution. In the UK, the Office of the Deputy Prime Minister (ODPM), with responsibility for local government, regional development, planning and some other functions on social issues http://www.odpm.gov.uk. The Ordnance Survey Great Britain (OSGB) is the national mapping agency for England, Scotland and Wales and a key institution in GII. In Poland, the main organization involved in GII is the Head Office of Geodesy and Cartography (GUGiK), but other institutions such as the Ministries of Construction and Environment also include GII in their operating procedures.

New pieces of legislation may be used to set up new GII institutions or functions. For example, as we saw in the last lecture, in Slovenia the "GeoInformation Centre" (GIC), was set up as the main GII in 1991. At its inception, the following functions were defined:

- Regulation and coordination of GI policy at a national level, and co-operation with other national and international organizations in GI-related standardization, legislation, policy, and legal and organizational aspects of data exchange and distribution,
- Development of user services including users requirements analysis, translation of requirements in terms of information processing, technical advice, linking information users and providers and quality support (preparation of quality manual, quality assurance, quality audits);
- Development of metadata services, remote access to metadata catalogues, and data provision through distributed data warehouse systems;
- Raising awareness of importance of an IT infrastructure, including human resources management, research and development, provision of tools, training, and data integration.

The data acquisition for this institution and others involved in GII in Slovenia (municipalities, 12 regional geodetic administrations, 46 government supervised branch offices and private companies) is controlled by the Geodetic Activities Act. The Spatial Planning Act also states that state and municipalities maintain a spatial data system that monitors the both the spatial planning and management. These laws ensure that institutions develop GII as a major or minor aspect of their mandated functions.

2.3 Human Resources.

Human resource issues, including training, skill acquisition and management are extremely important for the development of GII. In this section we look at models for GIS training. This includes the use of hardware and software, the management skills of the instructors, and the learning capabilities of the trainees. An important point in these models is the procedure of training, which allows repetitive trials to achieve the desired result.

2.3.1 Human issues with GIS.

These involve:

- Managerial issues which involve the authority, staff levels, management of resources, human resources development, leadership, strategies for private sector involvement, and other management and system support issues. When a new GII is developed, both pre-existing management systems and new methods may be developed. In the UK examples considered, new organisations were set up (for example the e-Government Unit) and pre-existing organisations were also used (the Ordnance Survey which converted to the production of digital products).
- Educational issues: The development of GII can only take place with the support of a strong knowledge and education infrastructure. Data producers needed to be trained to use the metadata standard. Trained human resources are needed for maintaining, upgrading, and using the system. The training of the trainers needed to be addressed in the early stage of implementation. There should be a short and long term training programs for managers, system administrators, and system operators, both at the infrastructure and organizational level. For example, in the Polish case, it was emphasised that the GII developments required the skills of highly trained specialists, obtained from a network of professionals.

2.3.2 Training and use of GII.

This has to do with optimising the usability of the GIS. GIS must be taught to people, either in formal education institutions, or on the job. This requires the design of a system that enables quick learning and retention of skills (Figure 1). An important issue is the mental model of learners, users and experts. The effectiveness of a GII may depend largely on the ability of people to learn, and hence manage, teach and develop, the required methods. Aspects of the mental model require of users include:

- Problem definition. Relevant real-world phenomena must be classified, linked into a problem or objective, to allow the rational design of a GIS, with execution and evaluation.
- Defining required system functions. The design of the system functionality is adapted or selected to resolve the problem.
- User interface design. The means by which the operator connects with the system, including the hardware and software.
- Visualisations design. This includes information held within the system, describing the real world phenomena and the results of the GIS analysis and modelling procedures which are displayed by the user through visual representations.
- Usability testing. Methods used to evaluate the ease with which the system can be used.

- Instruction and testing. The formal and informal help required by the user for the efficient operation of the system.

Usability testing requires assessment of the effectiveness of training and the assimilation by users of the methods used. These may be planned according to the following points:

- Targets to be achieved.
- The content.
- The order of presentation.
- The speed at which it is presented.
- The opportunity for repetition.
- The form of practice to be provided.
- Methods of testing.

The diagrams below show the stages of learning processes that must be considered in institutions.

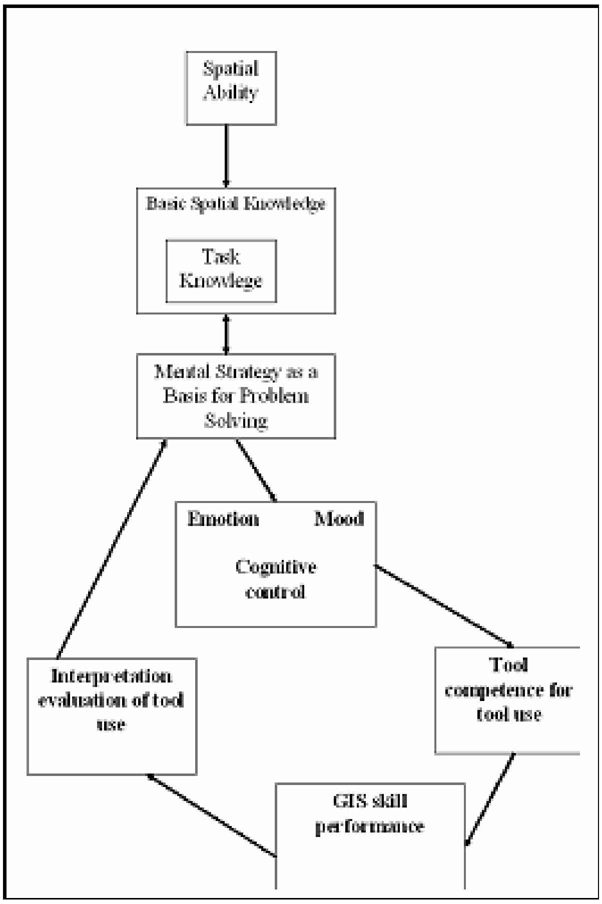

Figure 1. In this diagram, the learning speed of the GIS user is defined by his/her spatial ability, qualified by the basic spatial knowledge and understanding of the task. The formulation of the mental strategy for problem solving uses this knowledge and, based on factors such as emotion, mood, cognitive control, tool competence and interpretation of the outcome, GIS skill performance may be evaluated

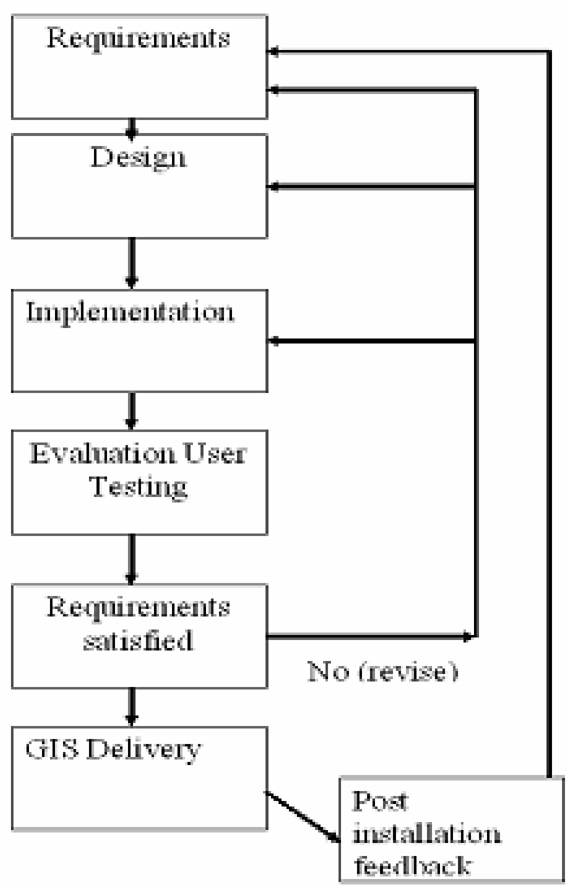

Figure 4. In this diagram, the implementation of a GII passes through several stages, which are all reliant of the organisational and human resources. Post installation feedback may be dependent not only on the problems within the institution, but also in the wider environment

2.4 Conclusion

In this module, we have looked at the institutional setting of GII. The institutional setting of a GII is a crucial aspect of GII development, and as we have seen in this section such a setting is related to all other aspects of GII development and management. The organizational, human resources and legal aspects of an institution are all related, and are also related to the wider environment of the institutional network, both nationally and internationally. The success of a GII, and the extent to which it may be expanded, is dependent on these environmental factors. Considering the earlier modules, we can see how the issues discussed in this module enable a comparison of the status of GII in different countries. We must consider the local technological, economic and political contexts in particular regions and countries. As in the last module, the results show that a well developed GII, in a good institutional setting is a benefit to any country.

Module Self-Study Questions

- If you joined a GII in an institution, newly trained in GIS techniques, what facilities would you expect? How do you think the environment outside the institution might affect the needs of a GIS trained person?
- Do you think that an organisation or institution can become too dependent on GIS? How would the institution ensure that GII is effective without becoming overly dependent on it?
- Do you think that geospatial data should be an open access resource? What restrictions would you place on public and private access to GII data and information?
- How would you improve the institutional setting of a GII? In what way are all the issues you would consider related?

Suggested Readings

- van Loenen V.B. and de Jong, J. 2006. Institutions Matter: the impact of institutional choices relative to access policy and data quality on the development of geographic information infrastructures. Website: http://www.gsdi9.cl/english/papers/TS8.2paper.pdf
- Groot, R. and McLaughlin, J. 2000. Geospatial Data Infrastructure. Oxford University Press, Oxford.
- Turk, A. The Relevance of Human Factors to Geographical Information Systems. In Medyckyj-Scott, D. and Hearnshaw, H.W. (1993) Human Factors in Geographical Information Systems. pp. 15-31.

Assignments

- Assignment 1: Institutional Framework for GII

3 Technical Framework for GII

In this module, we will examine the hardware, network components and software that make a Geographic Information Infrastructure possible. When most people think of GIS and GII, what comes first to mind is computer hardware. While it is true that this is important, and that GII would not be possible without it, computer hardware is really just the means to an end. Computer components make critical tasks of creating, documenting and sharing spatial information possible. Ultimately, it is the human contribution – creativity and organizations that permit GII to exist at all.

In Lecture 1, we will look at the hardware components of GII, the computers, storage devices and data backup technologies that make GII possible. Although we will briefly describe the hardware components for Geographical Information Systems, we will not focus on these, since they are somewhat peripheral to the whole discussion of Geographic Information Infrastructure. Lecture 2 will focus on the network infrastructure of GII. Without the technology to connect computers, there would be no GII. However, with the development of networks and the increase in value in what is transmitted on those networks, criminal elements have been attracted to the combination of accessibility and value. Thus, the second part of Lecture 4 discusses network security and techniques to ensure that only authorised people have access to the Lithuanian GII. Lecture 5 of this module discusses software components for GII, including technologies for presenting data on the Internet (portals), financial transaction software and supporting GIS software.

This module is the basis for a more “high level” discussion in Module 4, which looks at how these components are connected to create the GII.

Outline:

9. GIS Hardware
10. GII Hardware
11. Preventing Data Loss
12. GII Network Components
13. GII Software Components
14. Conclusion

3.1 GIS Hardware

Before we discuss the computer components of GII directly, let us examine the components of Geographical Information Systems. GIS computer hardware is a small part of the total GIS, which includes other elements such as software, data, methodology, people and applications.

We can divide GIS hardware into three classes: input devices, processing devices (computers) and output devices.

3.1.1 Input Devices

In recent years, with the development of the GIS industry leading to an increase in the amount of data that is available for GIS analysis, input devices have become less important than they once were. Many educational institutions teach very little in the way of digitising and scanning, instead placing emphasis on the analysis of spatial data.

For years, one of the most important pieces of GIS hardware was the *digitising tablet*, a table-like device on which maps could be placed (Figure 1). By following lines on the map with a device called a puck (also known as a cursor), a GIS operator could digitise the data, thus converting it into digital format for use with a GIS.

Figure 1. The GTCO Surface-Lit AccuTab Digitising Tablet (Source: http://www.gtcocalcomp.com/photos/PHatsl500.jpg)

Digitising existing maps has a number of inherent limitations, and this is one reason why it has declined in popularity in recent years.

1. Digitising is an inherently slow process. It may take hours or days to digitise a map.
2. Digitising is inherently error prone. Although some people are very good at following lines for hours, it is a tedious process. After several hours of digitising, the chances of human error increase.
3. Although digitisers are reasonably accurate at reporting the position of the puck (digitising device) on the tablet (~0.1 mm), the actual location recorded may be meters away from its correct position; this problem becomes greater at small map scales.
4. The data on the original map may have been displaced from its correct planimetric position to make the map more readable.

A scanner is a device which solved three out of the above four problems, but which creates a few of its own. Rather than relying on a human operator to enter all of the linework on a map, a scanner is an optomechanical device that makes a digital copy of a paper map. Because a human is not involved with the entry of every single point into the computer, the overall results for a scanned map are more consistent than for a digitised map. The results are faster, and are as accurate as digitisation. Automatic tools help to convert the raw raster input from the scanner into vector data.

Once again, the results of scanning a paper map are dependent on the original quality of the paper map. Both digitised and scanned data must be cleaned up before they can be used in a GIS, which also slows down the entry of new data into a GIS. Generalization and cartographic displacement make the paper map easier to read, but they distort the positional accuracy of features. Scanning also requires a very high quality copy of the paper map, preferably fresh from the printing press. Although the digitising process is less susceptible to poor quality maps than scanning, local distortions are difficult to remove from digitised maps.

3.1.2 Processing Devices

Geographic Information Systems have benefited greatly by the enormous increase in computer capabilities over the past thirty years. Early GISs were available only to large governments, but as computer hardware capabilities have increased, GIS capabilities are becoming mainstream. Google Earth is an example of a GIS-based product that was not even possible a decade ago, but which is now available for free to any owner of a moderately powered Personal Computer with an Internet connection.

Computers that could run GIS software used to be expensive to build; even when processors became powerful enough to permit a GIS to be run on a Personal Computer in the 1990's, these computers needed to be configured with high-end graphics cards and large amounts of memory to be useful as a GIS workstation. Today, even middle-of-the-road laptop computers have sufficient power to perform moderately sophisticated GIS analysis without modification (Figure 2). Certainly, if a computer is being used for GIS analysis on a permanent basis, it makes sense to augment the capability of the workstation. Adding a second (or third) monitor, a quality keyboard and mouse, and a connection to a file server to enable large files to be manipulated and stored, can be helpful, but these days, such items are more in the realm of "convenience" rather than "necessity."

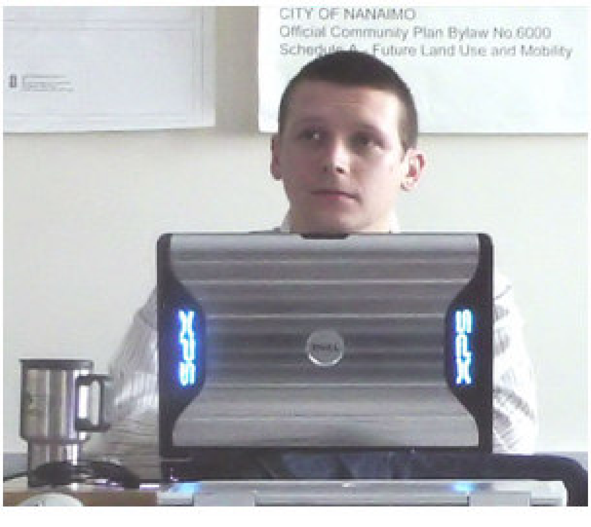

Figure 2. Laptop computers are now powerful enough for GIS analysis

3.1.3 Output Devices

Just as input devices had greater importance in the past than they do now, because of the change from a focus on data creation to data processing, this shift has also reduced the importance of output devices. Whereas people used to look at GIS in terms of its ability to create maps, today it is viewed as a tool to generate information. In many cases, that information is being submitted to a GII for use by the public or other professionals.

The most common output device for a GIS uses to be a *plotter*, which was a device for printing maps on large sheets of paper. In the past, plotters made use of pens driven by a robotic arm, or by ink, which was attracted to the paper by an electrostatic charge. Pen plotters were expensive devices that could create maps with up to eight colours, but had difficulties creating areas of solid colour. Electrostatic plotters could produce filled areas very easily, but they were extremely expensive, which limited their adoption. The development of inkjet technology allowed manufacturers to build inexpensive large format colour plotters, and this is why they have dominated the plotter market for over a decade now (Figure 3). Other than using larger ink reservoirs and having feed mechanisms for large rolls of paper, colour inkjet plotters use virtually the same technology as desktop inkjet printers.

In the early 1980's a GIS would have consisted of a standalone or networked computer connected with digitisers, scanners and plotters. With the development of the GIS industry, however, the value of the network has increased, whereas the value of the input and output devices has declined. Indeed, it is now possible never to use these devices – data is downloaded from the Internet, processed using a GIS and is then made available. The development of Geographic Information Infrastructures promises to strengthen this trend, and to make the GIS not just a tool for the specialist, but also a tool for everybody else.

Figure 3. Hewlett-Packard 820 MFP Scanner/Plotter combination . This is a modernised version of technology that was introduced over 15 years ago (http://h71016.www7.hp.com/ctoBases.asp?oi=E9CED&BEID=19701&SBLID=&ProductLineId=503&FamilyId=2367&LowBaseId=7088&LowPrice=$1,288.00#)

3.2 GII Hardware

3.2.1 Servers

If the personal computer has become the dominant class of computer for running GIS, then server computers are the dominant type of computer used to build Geographic Information Infrastructures. These computers are designed to supply data and applications to hundreds of users over the Internet. These are the same class of computers that drive Internet Service Providers, e-Commerce vendors and website publishers. Although these machines are reasonably fast by today's standards, they do not compete with larger mainframes and supercomputers used in scientific and business computing applications.

The key strength of servers is reliability. These machines are built with industrial strength components such as faster processors and memory, greater amounts of memory, redundancy in parts such as disc drives and power supplies that fail most often, and very fast network connections. Servers are designed to operate continuously without human intervention. Because server computers are not typically used for day to day work, as PC's are, pieces of hardware and parts of the operating system that are not required are not used on servers. Thus, it is common to find dozens of servers mounted in racks in a dark, air-conditioned room with no keyboards or monitors (Figure 4). If you look inside, you may find that these machines do not even have graphics cards, or the ability to display graphics in their operating systems. These machines are designed to be run entirely by remote control from another computer on the network.

Another key distinguishing characteristic of a server is that it is designed to be modular. Servers are built to connect easily with other computers by local area networks or via the Internet, so if a machine becomes overloaded, a new machine can be added. When server load grows over a long periods of time, new servers can be added to support the additional load. When server load fluctuates, such as for a web site that is very busy at certain times of the day, servers may be added dynamically though a load balancing server, which distributes incoming requests for information to different servers.

When many servers are required, they are commonly assigned a particular piece of software to run. Thus, we may encounter a "web server" or a "database server" in a larger operation. These servers are connected together to create an entire web portal, such as will be used for the Lithuanian GII when it is completed.

Figure 4. Rack Mounted Server Computers. Image: JohnSeb. Licensed under Creative Commons Attribution-Share Alike 2.0 Generic

3.2.2 Storage

When most people think of computer storage, they think of a hard disc humming away inside of a computer. However, there is much more to storage than just hard discs.

Computers are wonderful devices, but they are very forgetful. Every time you restart your computer, you must reload the operating system, any programs that you are using and the data that was being processed by these programs. Fortunately, computer engineers have developed storage techniques that enable operating systems, programs and data to be stored for long periods. There are three fundamental technologies that are used to store information when the computer is turned off: Non-Volatile Solid State Memory, Optical Media and Magnetic Media. We will begin with the newest of the technologies so that we can discuss magnetic media in detail.

Non-Volatile Solid State Memory

Recent demands from the consumer electronics industry for lighter, more shock resistant computer memory have led to the rapid development of non-volatile solid-state memory. This type of memory, known in various forms as Flash ROM, Compact Flash cards, Secure Digital cards, Memory Sticks, and Thumb Drives, store data in a form of computer memory that is not erased when the power supply is removed, unlike conventional computer memory.

Because of memory-hungry consumer electronics, particularly digital music players and digital cameras, the cost of this type of memory has been declining rapidly. It is now possible to have more than 1 GB stored on a card not much bigger than a coin (Figure 5). Non-volatile solid-state memory has a number of important advantages over traditional hard discs, including being very fast, very durable (great shock resistance and no moving parts), and using very little power. The costs of such media are still quite high (still about 30x as much as conventional hard discs), but recently, solid-state hard disc replacements for laptop computers have been introduced to the marketplace, and will soon become competitive with hard discs.

Figure 5. Non-volatile solid-state memory devices. Lexar 2 GB "thumb drive" (left) and SanDisc 512 MB Secure Digital Card (right). Pen is shown for scale.

Thumb drives, which contain non-volatile solid-state memory attached to a USB connector, have already taken the place of floppy discs, because they allow very large files to be transferred between computers (Khurshudov, 2001).

Optical Media

Optical media have also acted to replace the floppy disc. Ever since Compact Disc (CD) burners became available in 1985, it has been much easier to write large files to an optical disc than it was to split large files onto multiple floppy discs. CDs and Digital Versatile Discs (DVDs, also known as Digital Video Discs) are also very inexpensive to produce and will last a long time, provided they are treated with care. Because optical discs are removable, it is easy to scratch the exposed media. While some minor scratching is inevitable and does no damage, a large scratch may refract the laser beam severely enough to render the CD or DVD unreadable.

CDs can store up to 650 MB on a single disc, and DVDs can store up to 4.7 GB on a single-sided disc.[1] Both types of disc share the same physical dimensions, which makes it possible to read and write CD-ROMs in a DVD drive. DVD-ROM burners are able to do this by using a laser with a shorter wavelength of red light to encode more information in a given amount of area. Two new competing high-density DVD formats are now available (BluRay and High Density DVD (HD-DVD)), which promise densities up to 50 GB per disc by using an ultra-short wavelength blue laser.

Although optical discs such as CDs and DVDs share many characteristics with hard disc, the lasers that reside on the read head in the disc are heavier than magnetic read heads for hard discs, and cannot accelerate as rapidly, resulting in lower overall access times for optical drives than for hard discs. What optical drives lack in speed, they make up for in reliability. Since the read/write heads in optical drives do not need to come as close to the disc as read/write heads do to hard discs, optical discs are more robust, and are less susceptible to damage from being bumped while rotating (Khurshudov, 2001).

Magnetic Media

Magnetic media dates from the 1950's, when core memory was used as the main memory of some of the earliest computers (Figure 6). Created as an alternative to vacuum tubes (transistors had not yet been invented), core memory could exist in a positive state (representing a 1) or a negative state (representing a 0). A similar technique could be applied to any material containing magnetic particles; the particles could be charged with one polarity to represent a 1 and with the opposite polarity to represent a 0. Although the magnetic particles could theoretically be placed on an object having any shape, two forms of magnetic media became dominant: discs and tapes.

[1] Technically, the DVD specification allows for up to 17.08 GB per disc using a double-sided, double-layer configuration, but these are not commonly available.

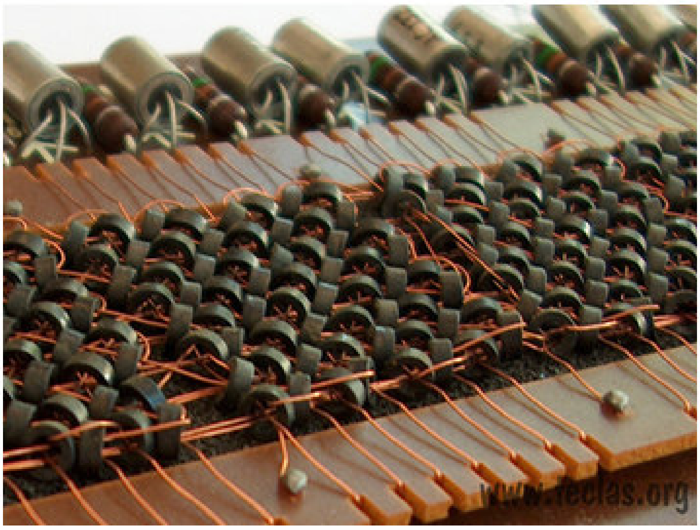

Figure 6. Magnetic core media from an Olympia RAE Desktop Calculator. Notice the torus-shaped magnetic cores, each of which can store 1 byte of data. Image: teclasorg. Licensed under Creative Commons Attribution-Share Alike 2.0 Generic.

The chief difference between a hard disc and a tape has to do with access speed and capacity. With a spinning disc, it is easy to access any part of the disc, with only a few microseconds' delay, because the disc spins at several thousand RPM, and the read/write head can move back and forth across the disc head rapidly. The trade-off for this method of rapid access is that the amount of surface area on a disc is quite low. To some extent, this can be solved by placing multiple discs on a single spindle, but even with this innovation, the amount of physical space available for storage is limited by the diameter of the disc.

As storage densities have increased, we have seen a decrease in disc size from 8 to 5.25, and then to 3.5-inch floppy disc formats, and from 5.25, to 3.5, 3.0, 2.5, and 1.8-inch hard discs. A wide variety of removable magnetic media drives have been developed, including improvements on floppy disc technology, and removable hard discs such as Zip, JAZ, ORB, and SuperDrive formats (Khurshudov, 2001).

Over the past 40 years, the recording density of magnetic media has been doubling every two years, (Khurshudov, 2001). This has led to a steady decrease in the price of storing each megabyte of memory. No matter how much the storage density increases, the small amount of surface area on hard discs always places them at a disadvantage to tape formats, which have much greater surface areas. This is why both forms of media have been around for decades, and will likely remain with us far into the future.

Media Lifetimes

Companies that produce magnetic, optical and solid-state memory must be able to compete effectively with other companies offering similar products. Fortunately, companies compete with one another based on the quality and expected lifetime of the media that they produce. This has led to rigorous quality standards and testing of all media types, magnetic and optical media in particular.

This is not to say that all media are equal, but it is important to recognise that most storage and backup media are designed to be used a great deal. There are differences in manufacturing quality, but the competition between media manufacturers tends to minimise the differences.

The (U.S.) National Media Laboratory, 3M Corporation and Carnegie-Mellon University assessed the chemical, thermal and archival stability of magnetic tapes, and found that most magnetic tape has a 15-30 year lifespan under normal usage. This assumes that the tapes are protected from environmental degradation, such as dust, high temperatures, high humidity, magnetic fields, corrosive chemicals and improper handling. In general, magnetic tapes need to be given 24 hours to adjust to changes in temperature and humidity (Chesher, 2003).

Another estimate suggests that high-density ½ inch tape cartridges have a lifespan of between 10 and 30 years when 10 GB per week are written to the tape (such as might be expected during incremental backups). Not surprisingly, the lifespans of these tape cartridges were found to be inversely proportional to the amount that they are used (Chesher, 2003).

Rewritable optical media have similar lifespans to magnetic tape. These discs (CD-RW, DVD-RW, DVD+RW) will last about 30 years (Chesher, 2003). Perhaps the most long-lived of the optical media are CD-R and DVD-R discs. These discs can be written once, and cannot be overwritten. Other than the standard drawbacks of optical media, such as susceptibility to damage from scratches as well as excessive light, humidity, and temperature, this type of optical disc is excellent for long-term data storage. It is estimated that a CD-R or DVD-R can last up to 100 years given proper storage conditions (Cougias, 2003).

Solid-state non-volatile memory, despite having no moving parts, does not last forever. The circuits in flash memory chips have a lifetime of only 100,000 to 300,000 cycles, so if a particular file was being constantly modified, for example if a database management system was writing to flash memory, the lifetime of the solid-state memory might not be as long as expected. In addition, with so many transistors on a single chip (thumb drives can now store up to 8 GB), it becomes quite likely that an individual transistor will eventually fail. Fortunately, the management software for such drives distributes the location of the writes across the chip so that no particular transistor gets overworked (Wikipedia "Solid-state drive", 2007).

One important factor to consider when thinking about long-term storage of data is that new technology may be more of a factor than media wear in determining how long a backup can be read. Although the physical format of tape cartridges may remain the same, changes in the format of the tapes themselves or the technology used to read and write data may prevent backwards compatibility (Chesher, 2003). System administrators charged with the task of maintaining data archives need to review storage technologies about every 10 years to ensure that the technology will still exist to read archival records in another 10 years. Such reviews may trigger a decision to transfer data records onto newer media.

Although this periodic replacement of media may be expensive, there are some advantages to be gained from the reduction in storage space required. Consider that about 1000 9-track tapes (40 MB each) can fit on one 40 GB DLT8000 tape cartridge (40 GB each). That would free up about 12 m^3 of space, or about a small room, and place all of the data on a single modern tape cartridge.

Market Factors

In addition to competition, there has also been a great demand from consumers and professionals for ever-larger capacity devices. Changes in the structure of the computer industry, such as the switch from large mainframe computers to smaller computers such as servers and personal computers, have led to a demand for small, high capacity disc drives. Continued miniaturization of computer components, which are now being driven by laptop computers, handheld computers, digital video cameras, personal video recorders and iPod music players, will continue into the future.

Another trend in the computer industry has been the move to larger, more capable programs with more refined user interfaces. This has increased the size of programs because of the additional computer code and graphics that are now part of every program. Of course, if you have a larger program, it means that you need additional storage capacity. Consider that DOS, the original PC operating system, shipped on a 1.44 MB floppy disc when it was introduced in the 1980's. Windows Vista, the latest successor to DOS, now ships on a 4.7 GB DVD, which can store more than 3000 times the amount of information!

Server reliability requirements have led to the development of drive arrays, in which information is stored redundantly in case one or more individual drives fails. These RAID (Redundant Array of Independent Drives) systems mean that more drives are required to store the same amount of information. It is possible to store data redundantly without having a complete duplicate, but the move to RAID systems has still increased the demand for high capacity disc drives (Khurshudov, 2001).

Storage Options for GII

The basis for all server storage is RAID technology. RAID systems have a number of advantages over single drives in terms of reliability. Rather than have a single large hard disc, a RAID system uses a number of inexpensive hard discs that are linked together as a single unit.

Data are split between individual drives, a technique that is known as *striping*. Striping leads to a faster read speed for the array than for individual discs, since only a small part of each file is read from each individual drive. With a single disc, the entire file must be read and the read speed is limited by the rotational speed of the disc (Figure 7).

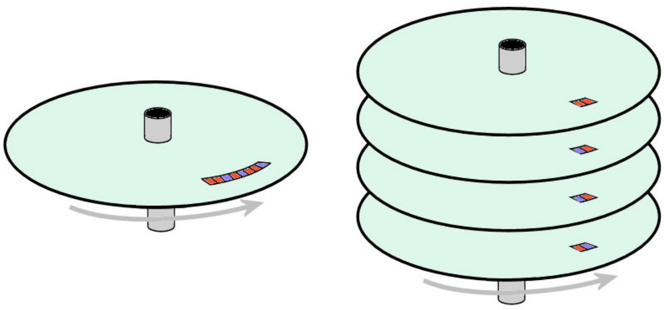

Figure 7. Striping of discs increases data access speed. If the rotational speeds of the discs are the same, the single disc at left will take 4 times as long to read 8 bits of data as reading the same data when it is striped on four discs

There are a number of different ways of splitting the data between drives, which provide different levels of protection. These methods are termed RAID-0 to RAID-6. The number is only an identifier for the type of data splitting used; there is no order to the methods used (Khurshudov, 2001). The most commonly used methods are RAID-0, 1, 3, and 5. RAID-2 requires custom disc drives, so is not available commercially, and RAID 4 is not commonly used because it offers few advantages over RAID-3 and 5 (Table 1). RAID-6 stores the parity twice, which provides the ability to lose a second drive while one drive is being rebuilt. This ability is important when large discs are being used in the RAID system, since it takes a long time to rebuild a large disc, which increases the risk that a second drive will fail during the rebuilding process.

Table 1. Characteristics of RAID Systems (Adapted from Khurshudov, 2001, p. 289)

RAID Level	Capacity	Large Data Transfers	Input/Output Rate	Data Availability
Single Disc	Fixed (100%)	Good	Good	10,000 to 1,000,000 Hours
RAID-0	Excellent	Very Good	Very Good	Poor*
RAID-1	Moderate (50%)	Good	Good	Good
RAID-2	Very Good	Good	Poor	Good
RAID-3	Very Good	Very Good	Poor	Good
RAID-4	Very Good	Very Good	Poor	Good
RAID-5	Very Good	Very Good	Good	Good
RAID-6	Very Good	Very Good	Good	Good
*Availability is equal to Mean Time Between Failure (MTBF) of one disc divided by the number of discs in the array. In other words, in a 4-disc array, the MTBF is ¼ of an individual disc.				

The RAID-0 method requires a minimum of 2 disc drives and merely provides striping, but no redundancy. This method provides fast data access and high data storage capacity relative to the other methods, but cannot automatically recover from a disc failure.

RAID-1 uses *data mirroring*, which is the most reliable of all RAID methods, since a complete copy of all data are stored in the array. RAID-1 is quite inefficient in terms of storage space, since every byte of data must be stored twice. It is, however, the only technique that is able to rebuild itself if 50% of the drives are lost.

It should be noted that data mirroring is a variation on the general technique of disc cloning. Mirroring ensures that what is written on one disc is also written on the other disc, almost immediately. Cloning also copies the entire contents of one disc to another, but runs as an backup procedure perhaps once a day, instead of continuously as in RAID-1.

RAID level 3 requires 5 individual disc drives and provides data redundancy without having to duplicate each piece of information as RAID-1 does. The key to this is to store a parity bit for each byte of data on a separate drive (also known as *Parity RAID*). A parity bit simply indicates whether an even or odd number of 1's occurs in a group of binary numbers. If any single binary number is erased, the parity bit can be used to determine what the value was. The parity bit enables the array to reconstruct the data if a single disc drive fails. If a data drive fails, then the parity bit can be used to determine the bit that was stored on the drive; if the drive storing the parity bits fails, then a complete set of data exists from which the parity data can be reconstructed (Figure 8).

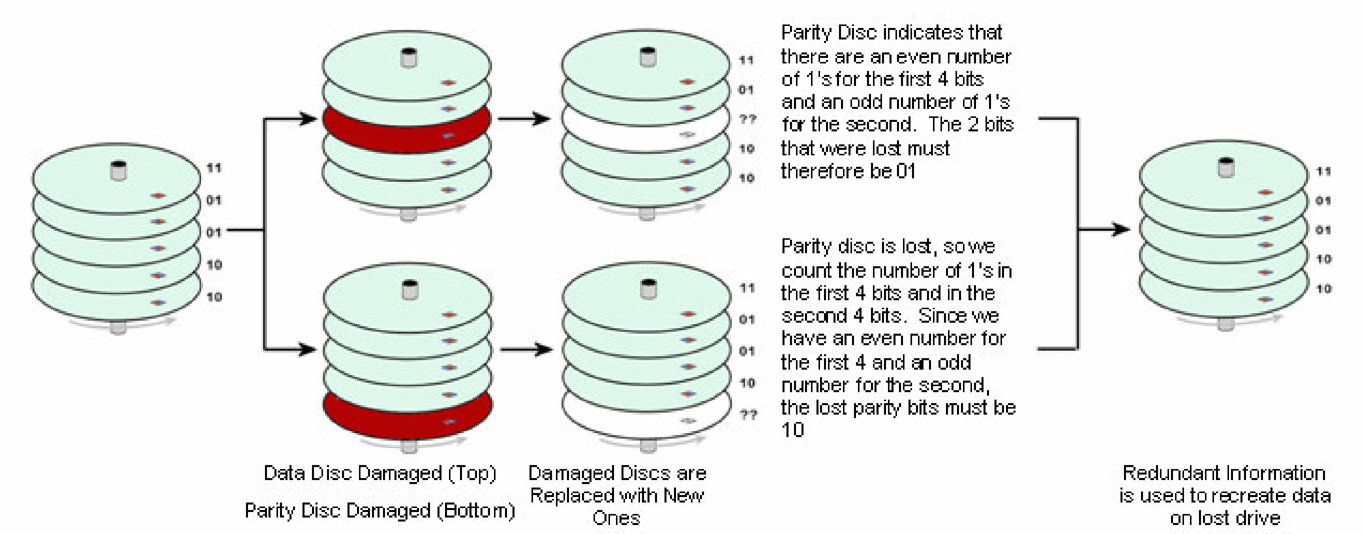

Figure 8. Recovery from a disc failure using RAID-3. If a data disc is lost (top), it can be recreated by analysing the data on the remaining data discs and the parity discs. If the parity disc is lost (bottom), it can be recreated by analysing the data discs.

RAID level 5 requires 4 individual disc drives, but instead of storing parity bits on a separate drive, the parity bits are duplicated, and a copy is stored on each disc.

The reason that we use RAID systems has to do with data redundancy. Normally, if a single disc drive fails, your only option to recover from the failure is to restore data from a backup tape. With a RAID system, drives are *hot swappable*, which means that a failed drive can be removed and replaced with a new unit while the RAID system continues to be used by the computer. Once a new drive is in place, the computer automatically rebuilds the content of the drive from the redundant information stored. Of course, during the reconstruction process, the performance of the disc degraded by up to 50%, but information is always accessible. A further enhancement of the idea of a hot swappable disc is the idea of a *hot spare*, in which a spare drive is stored in the RAID system, and can be called into use immediately if a drive fails. If a RAID array has a backup power supply as well as a hot spare, the two most common causes of storage system failure are eliminated, which enables RAID systems to run continuously over long periods of time.

Of all the RAID specifications, RAID-5 provides the best combination of security and economy. RAID-5 was originally created in 1978 by Norman Ken Ouchi at IBM, and was awarded US patent number 4,092,732. RAID-5 allows the data on the array to be recovered if a single disc drive fails. With the addition of a hot spare drive, RAID-5 arrays can be left unattended for many months with very low probability of data loss when fitted with a hot spare drive. Should a drive fail, the hot spare will be activated and the redundant information will automatically be copied to the hot spare while the array remains in operation (with degraded performance). At this point, the system administrator will replace the failed drive, which will become the next hot spare (Figure 9).

Figure 9. The Norco Technologies DS-500 5-Bay Hot Swappable SATA RAID Storage Array, which contains a hot spare. Notice the levers on each drive face that allow the rapid removal of each drive (Source: http://www.norcotek.com/item_detail.php?categoryid=8&modelno=ds-500)

In a RAID-5 array, the parity bits are distributed among all of the drives in the array (Figure 10).

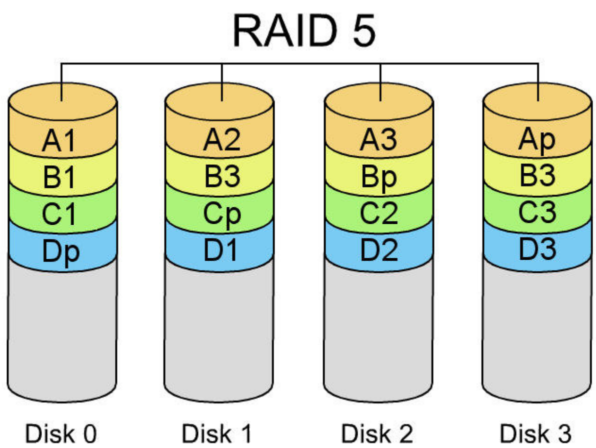

Figure 10. Schematic of RAID-5 array with four discs. Parity bits (Ap, Bp,Cp, Dp) are spread among the four discs for increased reliability. (Source: http://en.wikipedia.org/wiki/Image:RAID_5.svg, licensed under GNU Free Documentation License)

The reliability of a RAID-5 can be calculated based on the Mean Time Between Failure (MTBF) and the Bit Error Rate (BER) of individual drives. The MTBF is the mean time that an individual disc drives will operate before it fails. The BER represents the probability that an individual bit will be retrieved from a hard disc incorrectly; this value is measured as a probability. These values are used to calculate the Mean Time to Data Loss (MTTDL), which is the mean time for the loss of data and the entire array. Since individual drives may fail without destroying the data in the array, this value is used instead of the MTBF in RAID arrays.

As an example, the Seagate Cheetah 10K .7 300 GB hard disc has an MTBF value of 1.40×10^6 hours and a BER value of $1 / 10^{14}$. In a 12 disc RAID array, the MTTDL value is 4.44×10^6 hours (about 506 years), or more than three times that of an individual drive (Treadway, 2005).

Storage Networks

Another adaptation that increases the reliability of server-based computing is to separate the RAID systems and other discs and tape media from the server. This offers a number of advantages including increasing reliability, reducing ownership costs, centralising data management, and increasing the speed of backups and applications, since backups can run from a backup server without affecting the speed of applications running on the application servers.

Network Attached Storage (NAS) allows a storage server containing a RAID system to be attached to a Local Area Network as a network device. All computers requiring use of the RAID drives can access them as a network volume, in the same way that they would be used if they were made available through a server. This allows workstations on the network to be built without hard discs entirely; they simply boot directly from the storage server. Unfortunately, this system relies on very fast networks, since all data flows though the network. When the additional overhead of converting all data into packets is considered, it quickly becomes apparent that NAS is not the best solution in high data-flow environments such as GIIs. Network delays caused by competing traffic mean that NAS is much slower than a traditional direct connection between the RAID system and a server.

Another solution is a Storage Area Network (SAN), which eliminates many of the problems with Network Attached Storage. A SAN is "a network whose primary purpose is the transfer of data between computer systems and storage elements" (Tate *et al.,* 2006, p. 2). The main difference between a SAN and NAS is that a SAN does not use the regular local area network to transfer data. Data transfer in a SAN occurs on a dedicated data network that is separate from the LAN. Whereas Network Attached Storage can be used to provide a central repository of data such as accessible by both servers and individual workstations, a Storage Area Network is designed to make high-performance storage available to the servers that require it most. The servers may then make this available through the LAN to individual workstations, allowing the server to operate very efficiently (Figure 11).

Rationale for SAN's

It may seem odd that one way to improve server access to storage media would be to disconnect the media from the server. Directly-connected disks and tapes offer high data transfer speeds and work work very well when only a few computers are involved, but when we begin to deal with large data centres, such as the one which will support the Lithuanian GII, such systems become difficult to administer. The more data involved, the more it makes sense to create a Storage Area Network.

When storage is connected directly to a computer, it is difficult to access and use effectively. For example, to back up a disc drive that is directly connected to a server, either a tape drive needs to be directly connected to the server, or the data from the disc drive needs to be transferred from one server across the network to another server, and stored on servers tape drive. In the first case, the server uses many resources to perform the backup procedure, and in the second case, both servers are not fully available and LAN bandwidth is being used. Part of the problem is a result of the vastly increased disc capacities that have become available in the past few years; Input/Output and LAN speeds have not kept pace with the vastly increased disc capacities, so it has become increasingly difficult to back up high-capacity discs using LANs.

Another problem with locally mounted discs is that they encourage organisations to purchase excess disc capacity, which can become very expensive when we are dealing with hundreds of computers. The reason for this is because, when discs are mounted locally, there must be sufficient reserves of disc storage on each single computer. If a computer runs out of disc space it crashes, so there must be always sufficient extra space on each individual computer to allow it to keep operating.

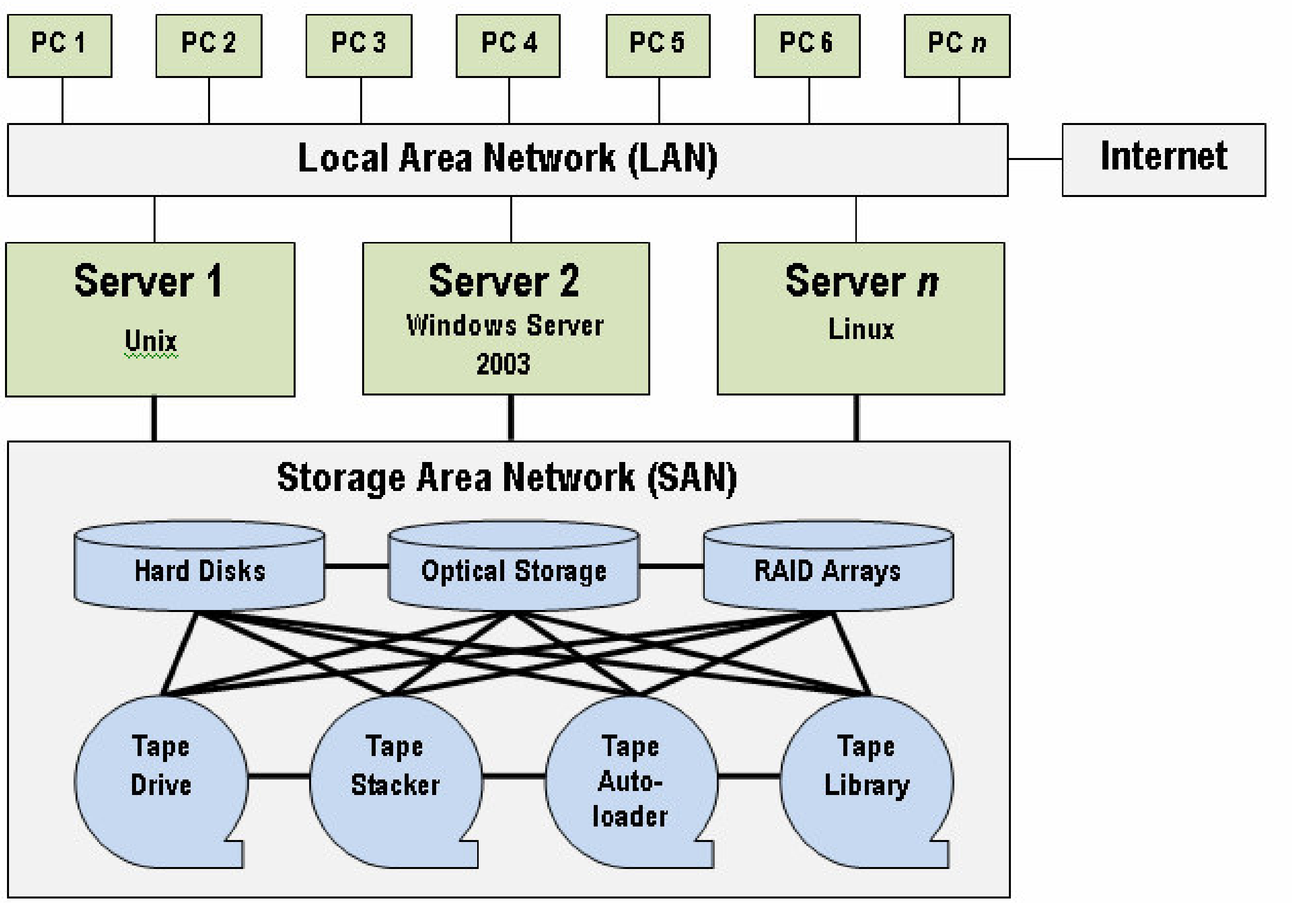

Figure 11. Storage Array Network attached to servers and to PC's via a Local Area Network (Source: based on Tate *et al.,* Figure 1-1, p. 3)

SAN's solve these problems by retaining the fast disc access speeds that are achieved by having local discs, while physically removing the discs from the server itself and placing them on a network where they can be accessed by all servers.

The configuration of a SAN offers the following advantages over locally mounted disc systems:

- SAN's allow each server to access unlimited amounts of disc storage
- SAN's allow an optimal amount of disc storage to be purchased, and disc usage is shared between servers
    - Servers may run any operating system, but can still access the discs on the SAN
    - Any server can access any storage device on the SAN
- SAN's allow all types of disc and tape storage (discs, RAID arrays, optical drives, tape drives (individual, stackers, autoloaders, libraries) to be combined and used more effectively
    - Backups can be performed rapidly, without server intervention, while discs are being accessed
    - Backups do not occupy bandwidth on the LAN because they are handled by the separate network in the SAN

- Storage processing is no longer handled by servers
  - Individual commands are issued to SAN and the network completes the task itself
  - Data can be backed up by network without using server processors
  - SAN can handle synchronous or asynchronous copying of files by itself
- Servers can be placed together in one room, and SAN can be placed in another room, allowing for more efficient administration and maintenance

Storage Area Networks offer one additional feature not available with Network Attached Storage. Since all servers are connected to the SAN, they are able to directly communicate using the SAN, rather than the much slower LAN (Tate *et al,* 2006)

Design of SAN's

Most servers make use of protocols such as the Small Computer Systems Interface (SCSI) or Serial Advanced Technology Attachment (Serial ATA or SATA) to connect hard discs or RAID arrays to servers. These protocols combine hardware and software, and assume that the hard discs or RAID arrays are in close proximity to the computer. When we create a SAN, these protocols are transferred across a network, allowing servers to control the discs in the SAN as if they were locally connected.

A number of different methods are available to connect disc storage to servers in a SAN. These methods fall into two broad categories; Ethernet-based networks and supercomputing-derived networks.

The Ethernet-based networks make use of these same technologies found in Local Area Networks to communicate between the servers and the hard disks in the SAN. The difference here is that the communication between servers and discs is handled on a separate network from the LAN, which avoids some of the bottlenecks that are encountered in Network Attached Storage. A number of different competing protocols exist, such as iSCSI, which transfers SCSI signals over TCP/IP, HyperSCSI, in which the SCSI signals are transferred using Internet Protocol, and ATA signals being transferred using Internet Protocol. These systems work relatively well, but the networks on which the signals are transmitted are quite slow by modern standards, and new options are available that make use of research that was performed to create supercomputers ("Storage Area Network," Wikipedia).

Supercomputing-derived networks include InfiniBand (IB) and Fibre Channel (FC). InfiniBand is a high speed, 2.5 Gb per second switched fabric communications link which is designed to be scalable. New versions of InfiniBand increase the transmission rate to 5 and 10 Gb per second. Part of the reason that these new technologies are so fast is because there is much less overhead involved in sending information through these networks. Each TCP/IP or IP packet has 192 bits of header for each 32 bits of data; InfiniBand uses the 8B/10B encoding system, in which there is only two bits of overhead for every eight bits of data transmitted ("Infiniband," Wikipedia). A Storage Area Network makes use of InfiniBand by transferring SCSI signals using this protocol in much the same manner that HyperSCSI piggybacks the SCSI signal on top of IP.

The other supercomputer-derived network is Fibre Channel, which is the most commonly used network technology used in SAN's. Fibre Channel allows for burst transmission speeds of more than 100 MB/s. Unfortunately, Fibre Channel is expensive technology, which limits its use to areas where high-bandwidth data transfers are critical (Khurshudov, 2001). Recent refinements to the Fibre Channel technology are helping to reduce the complexity and expense of this technology.

Fibre Channel was originally developed in 1988, and approved as an ANSI standard in 1994. Despite its name, Fibre Channel does not necessarily use fibre-optics. Fibre Channel may also make use of twisted pair copper wires in addition to fibre-optic cables. Three different network topologies are possible using Fibre Channel, the first of which is Point-to-Point (FC-P2P), in which two devices are connected directly. A more flexible option is Fibre Channel Arbitrated Loop (FC-AL), in which devices are connected in a loop. FC-AL allow up to 128 devices to be connected. Fibre Channel Switched Fabric (FC-SW), the third option, requires all devices to be connected to Fibre Channel Switches. This configuration allows for 2^{24} possible connections. Because of the way that FC-SW is designed, every time a new device is added, the amount of bandwidth is increased proportionately, so the amount of bandwidth always scales with the number of machines that use it.

Fibre Channel uses an even more efficient encoding method than 8B/10B encoding system used in InfiniBand, which is the 64B/66B system, in which 66 bits of raw data are required to transfer 64 bits of information (“Fibre Channel,” Wikipedia).

Maintenance of SAN’s

The separation of servers from storage in a Storage Area Network reduces the overall maintenance requirements for both servers and hard disks. If a hard disk needs to be added to the SAN, it can simply be added and brought online, without having to shut down any servers. If a hard disk fails, it can simply be removed.

Similarly, if a server needs to be removed for maintenance, it can simply be replaced by another server in a matter of seconds. Because the servers boot directly from the SAN, changing a server is as easy as unplugging one computer and plugging in another. Depending on the amount of load on the GII, servers can be activated to take on increased load, or can be shut down if the load can be handled by fewer servers.

Because the tape backup units are part of the SAN, the tape backup can occur at any time, without having to remove the drive being backed up from service. Backups are also more efficient than in Network Attached Storage, because the tape drives have block-level access to files in a SAN, which allows only the changed parts of files to be backed up, rather than the entire file.

3.2.3 Power Supply

One of the key components in the facility in which the Lithuanian GII is hosted is ensuring that the computers have Uninterruptable Power Supplies (UPS) and power conditioners that filter out power spikes. Supplementary power may be provided by batteries that run just the computers and other hardware, or a generator that supplies power to the entire building in which the computers are housed. In addition to having a key piece of government infrastructure unavailable, outages or uncontrolled changes in the power supply can significantly shorten the lives of electronic components.

3.3 Preventing Data Loss

Although RAID and Uninterruptable Power Systems greatly help to prevent data loss, they are only parts of a much larger set of strategies. There are numerous ways to lose data, and, unfortunately, there is no sure-fire solution to this problem. The best strategy is to have a solid system in place consisting of multiple, overlapping procedures to prevent data loss.

The system begins with what might be called “data hygiene.” Having an environment that is suitable for data processing contributes greatly to the success of an IT operation. Simple tasks like keeping computer equipment in a clean environment, ensuring that media are placed in their proper storage cases when not in use, and avoiding magnetic fields and electrostatic discharge, both of which can be hazardous to electronic components, go a long way to preventing many different kinds of data loss. Ensuring that discs are placed in a safe environment where they are unlikely to be jarred or shaken, in a climate controlled room helps to prevent the failure of media and electronic components due to overheating (Khurshudov, 2001).

Aside from common sense rules for treating your backup media with care, there are a number of different strategies that can be pursued to help prevent the loss of data. The first of these is duplication – making a complete copy of important files. This is relatively easy to do if you have a CD-ROM or DVD burner or even a backup tape drive on your computer. Network utilities can also be devised to make duplicates of files with a single command. Replication is a little more advanced than duplication, in that a copy is made dynamically, so that even changes made to a file in the last seconds before a hard disc failure will be archived on another disc. Along similar lines, Versioning never allows a file to be overwritten, but instead always creates a new file with a version number appended to it.

A very effective way of preventing the loss of data is to invest in “fault tolerant” systems. We know that humans make mistakes, so we should design an environment in which people can make mistakes, but are able to correct their mistakes without too much difficulty. Hard discs are going to fail eventually, so if we plan for them to fail, then the result need not be catastrophic.

Khurshudov, 2001 discusses the types of human errors that cause of data loss (Figure 12). By far the greatest risk is accidental deletion of data. Another risk is the intentional deletion of data (i.e. deleting the wrong file). The remaining human cause is the impact of computer viruses, and their potential to delete or corrupt files. Here, the best solution is to use effective antivirus software to prevent viruses from infecting your network (Khurshudov, 2001).

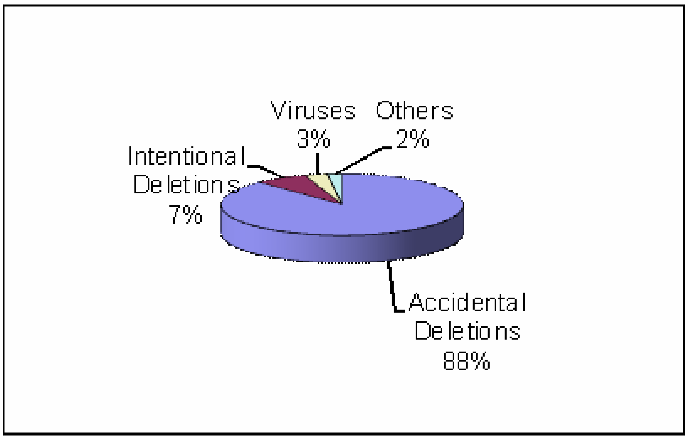

Figure 12. Human causes for computer data loss (Khurshudov, 2001)

One way to reduce the need for restoring files from tape backup is to make an “Undelete” command available to all employees (Cougias, 2003). While the Recycle Bin on Windows systems helps to reduce the need to recover files from tape, it is still possible to delete files completely, particularly on network drives. When working on network drives, files are not transferred to the Recycle Bin when they are deleted, so there is no copy of the file to restore. An Undelete command takes advantage of the fact that when a file is deleted, only the entry for the file in the disc catalogue is deleted, not the data itself. An undelete program can scan the disc for files that are not recorded in the disc directory, and allows the disc directory entry to be recreated, in effect bringing back a file from the dead. One limitation to this technique is that it will work as long as new data has not overwritten the old file information on the disc. For this reason, Undelete commands work best when the user immediately recognises that a file was accidentally deleted and sets about undeleting it. The longer the user waits, the less likely it is that the file can be recovered.

In addition to the Undelete command, users should be able to make backups of important files easily and quickly. Cougias (2003) recommends setting up an easy-to-use network based backup system, so that the user can make backup copies of important files as soon as particular milestones are reached. If, for example, the user has just completed a time-consuming piece of GIS analysis, easily being able to make a network-based backup encourages the user to be self-reliant. If the user accidentally deletes or corrupts their file after the backup has been made, then the restore command should be available to the user, and it should be as simple to use as the backup command. This presents a quick and effective alternative to backing up important files to a Flash ROM drive, floppy disc drive, DVD or CD, and eliminates the chances that removable media will be misplaced or lost (Cougias, 2003).

A slight variation on this theme is to make two copies of the file when a user requests a backup. One copy can be restored, and the other is permanently saved with a version number appended to the file name. This means that if the user archives a file, then corrupts the file and archives it again without realising that it is corrupted, the original archive file is still available to the system administrator.

Archiving is the creation of a permanent backup file, preferably on multiple copies of very stable media, when a project ends, or comes to a lull when creating a permanent record makes sense.

The other major method for avoiding data loss is to make backup copies of your data. As we discussed in Section 0, magnetic tape allows vast quantities of information to be stored, although it is difficult to access individual files because the tape needs to be moved to the correct position to read particular pieces of data. Magnetic tape however is perfect for backing up hard discs, where the entire contents of the hard disc can be transferred to magnetic tape during a backup session. If a disc fails, then the contents of the entire tape can be copied back onto a new disc. If a single file is erased, then the file can be found on the tape for restoration to disc. Both restore operations may take several hours to complete, but this is acceptable, because such disasters occur relatively infrequently.

In many offices, computers are relatively idle during the night, and we can use this downtime to allow scheduled backups to take place. Using such a system, each day’s work is automatically backed up by the computer, so at most, an entire day’s work is lost.

Special back up techniques are required for systems such as Database Management Systems, which rely on files always being open. These transaction management commands allow changes to be “rolled back” if the result of a command is undesirable (Cougias, 2003). In a highly dynamic environment such as this, where disc access is occurring constantly, disc mirroring is an important strategy for preserving data integrity.

3.3.1 Backup Tape Technology

Because of the physical differences between hard discs and tapes, they are used for different purposes: discs when moderate quantities of information must be accessed rapidly and tapes when truly enormous amounts of data need to be accessed more slowly. Tapes can have a much greater amount of surface area than a disc, which allows them to have much greater storage capacity overall. This greater capacity comes at a price, however, in that it takes a long time to wind through a length of tape to access a particular file. This is the reason why there is a hard disc in every computer, and tape drives only in those computers that are used to back up information.

Many different variations on these two fundamental media formats have been created. The original reel to reel tapes used in the 1960’s and 1970’s gave way to a wide variety of tape cartridges in the 1980’s, a situation that is still with us today, as different manufacturers compete for a share of the backup media market.

Magnetic tape comes in many different widths and cartridge formats. Tape may come in ½, ¼ inch as well as 8mm and 4mm tape widths. Various cartridge formats can be used, such as 4mm Digital Audio Tape (DAT), Quarter-inch Cartridge (QIC), 8mm tape cartridges, DLT and SuperDLT cartridges (Figure 13) and Linear Tape Open (LTO) format (Khurshudov, 2001). Tapes may be read by a rotating head that scans across the tape (helical scan), or by a fixed head, which makes use of tape movement to read the tape. The following table (Table 2) describes some of the characteristics of current tape media.

Figure 13. SuperDLT Tape Cartridge. Image: redjar. Licensed under Creative Commons Attribution-Share Alike 2.0 Generic.

Table 2. Modern tape formats, scan methods, capacities, cost, and data transfer rates (Cougias, 2003)

Format	Type of Scan	Capacity Per Tape	Cost per GB ($ US)	Data Transfer Rate
AIT (AIT-1)	Helical	100 GB	$0.19	12 Mbps
DAT (DDS-4)	Helical	20 GB	$0.08	3 Mbps
DLT8000	Linear	40 GB	$0.14	6 Mbps
SuperDLT	Linear	110 GB	$0.14	11 Mbps
DLT1	Linear	40 GB	$0.14	3 Mbps
LTO (Ultrium)	Linear	100 GB	$0.07	16 Mbps
Mammoth/VXA	Helical	60 GB	$0.15	12 Mbps

A recurring theme seems to be the reintroduction of the same cartridge type about every 7-10 years with vastly improved magnetic tape inside. These “new” cartridges can be expected to store 10-100 times the amount of data that could be stored by the previous generation of cartridges.

Automated Tape Systems

With every increase in the capacity of individual tapes we see a corresponding increase in the capacity of hard discs. For this reason, there are always situations where individual tape cartridges are unable to store sufficient data. For example, it might be desirable to back up all of the servers on a Local Area Network in a single backup operation. To do this, the backup must be spread across more than one tape cartridge, since the total amount of data to be backed up exceeds the capacity of a single cartridge.

In larger backup situations such as these, it becomes difficult to track individual tape cartridges, so the likelihood of human error becomes greater.

The answer to these problems involves automated tape systems, which are essentially robotic devices to feed multiple tapes into a tape drive in a systematic fashion so that the backup can be left unattended. There are three levels of automated tape system, but we will discuss these from the simplest to the most complex.

A Tape Stacker is a simple device that allows a stack of tapes to be loaded into a tape drive sequentially, such that as one tape fills up, it is replaced by the next empty tape, and the process continues until the backup job is completed (or the system runs out of blank tapes). Although Tape Stackers are inexpensive, they are somewhat inflexible when the time comes to restore data, since a human operator must insert the correct tape from the stack in order to recover data.

To resolve this problem, Tape Autoloaders were created. These are essentially robots that can pick any tape out of a cartridge and load it into a tape drive. This allows both the backup and restore processes to run completely unattended. It also allows the tape racks in which the individual cartridges are loaded to be treated as individual units, so a single rack can be loaded and unloaded from the Tape Autoloader very rapidly.

Although Tape Autoloaders are quite flexible, they are limited by the size of the rack into which tape cartridges are loaded. Tape Libraries allow unlimited numbers of tape cartridges to be used. To keep track of the tape cartridges, each cartridge is bar-coded, and it may be loaded into one of many tape drives that are available in the Tape Library. This combination of large numbers of tapes and many tape drives allow Tape Libraries to offer large scale backup capabilities with near-online access add user-initiated file recovery using network-based file recovery software that controls the Tape Library (Khurshudov, 2001).

3.3.2 Backup Regimes

In an ideal world, a complete copy of all data would be made automatically every evening. Such a scenario of full backups means that there would be a complete record of the state of every file taken every 24 hours. However, the cost and time required to perform full backups on such a scale is prohibitive.

For this reason, daily backups consist most often of incremental backups, which store only those files that changed since the last full or incremental backup. Although incremental backups take much less time to execute and use less storage media, they are more difficult to use, because locating the correct version of a particular file may involve several hours of searching through multiple backup sets.

Most system administrators plan a backup regime that combines both full and incremental backups in an attempt to minimise the total costs of backup hardware, backup media and hourly wages involved in administering the whole setup.

Full backups, also known as recycle backups, involve completely erasing what was on the backup media previously, and making a complete copy of all the files on a file system. If the backup is made to a hard disc instead of tape media, this operation is the same as disc cloning, which was discussed earlier. Full backups should be run before and after dangerous or important events, such as the installation of new software, changes to a file system, or the departure of an employee (Cougias, 2003). Incremental backups are also known as normal, changed file, or evolutionary backups. These make use of a *file archive bit*, which is a binary flag that is stored with other file information such as the filename and file size. The file archive bit that is set by the operating system whenever a file is modified or created, and when it is backed up by the backup utility, the bit is changed back to indicate that the file is backed up.

Differential Backups are sometimes added to the mix to allow the time between full backups to be extended. A differential backup stores all changes that were made since the last full backup. They essentially duplicate several incremental backups, and allow a restore to be made from a single backup set rather than multiple incremental backups, but suffer from some of the same problems as full backups, since they consume a great deal of media.

It is important to recognise that a properly executed backup regime is very hard on storage media. Tapes may be overwritten several dozen times per year, depending on the regime used. Furthermore, the measure of media life, Mean Time Between Failure (MTBF) is a statistical measure (with a large standard deviation), so some tapes will last much longer than the posted MTBF value and some will last a very short period. With the large number of tapes that are involved in a backup regime, it is inevitable that a tape will eventually fail; however, by rotating tapes so that each tape is used evenly, the odds of an individual tape failure are reduced.

The actual schedule of full, incremental and differential backups is up to the system administrator. However, a system that rotates a set of tapes, keeps all but the current day's tapes offsite, and provides an orderly set of full weekly backups, with an appropriate number of full backup sets being preserved to provide a "memory" of what happened reduces the risk of catastrophe significantly.

Much of the discussion so far has focussed on the impact of a power surge, or a disc failure. However, there is a small possibility of a more significant event such as a fire. In a case like this, some planning about how to continue the operations of the company at another site, using borrowed equipment if necessary is invaluable. Not only should disc backups be kept offsite, but also critical business information, plans for the computer network, and a catalogue of all files all need to be stored offsite. It is one thing to restore a failed disc system and quite another to have to rebuild a computer network from scratch, and then restore the data from backups to a series of new computers. The September 11, 2001 attacks on the World Trade Center in New York City illustrated that a significant disaster (terrorist attack) can bankrupt companies if they do not have a comprehensive emergency plan. Those companies that had effective backup and contingency plans were able to shift operations to other offices, in some cases immediately.

Increasingly in our networked world, it is possible to back up significant amounts of data to an offsite location automatically. Such a system of online storage requires a high bandwidth Internet connection. The data is stored confidentially by companies that invest large amounts of money in having the best storage techniques, highly experienced staff and dedicated storage facilities. For the small proportion of files that are critical, it makes sense to keep multiple copies, and online storage facilities can help to ensure that this data is secure.

3.4 GII Network Components

A computer network is an interconnection of several computers. Network components include the hardware, software, physical links and data of a network. Networks are classified by the operating network layer and according to basic reference models considered as standards in the computer industry.

3.4.1 Network Classifications

Computer networks may be classified in different ways.

- By scale. These include, from smallest to largest: Personal Area Network (PAN), Local Area Network (LAN, usually 1 km), Campus Area Network (CAN), Metropolitan Area Network (MAN, usually city-wide), or Wide area network (WAN). Because Ethernet is becoming the standard network interface, end users are less concerned about these classifications than network administrators are. This is because network administrators sometimes adjust the network, to increase the Quality of Service (QoS).
- By the connection method. This is based on the hardware technology that is used to connect the individual devices in the network such as Ethernet, Wireless LAN or Power line communication. The Ethernet uses physical wiring for connection, using hubs, switches, bridges, and routers. With wireless LAN technology, the devices are connected without wiring, using radio frequency signals. Power line communication transmits network signals over the same copper wires that supply electricity to homes and businesses.
- By the functional relationships (these may be termed Network Architectures) that link network elements. Examples are Active Networking, Client-server and Peer-to-peer networking (workgroup) architectures.
- By the network topology. Network Topology signifies the way in which logical relations between networks are transmitted between components. This is independent of the physical configuration of the network. Some examples are Bus, Star, Ring, Mesh, Star-bus, Tree or Hierarchical topology and Complete Graph networks.
- By protocol, referring to the communications protocol that is being used on the network.

3.4.2 Network Hardware

Network Hardware refers to the physical materials that compose a network. These enable the operation of the networks described above. In addition to computers and connecting cables or fibre-optics, modems, routers and firewalls are hardware components that make Ethernet networks possible.

Modems

The word “modem” is a contraction of Modulator/Demodulator. Originally, modems were built to allow digital data to be transferred over analogue voice phone lines. Digital information would be modulated (converted to an audio signal) on one end of the connection and demodulated (converted back into a digital signal) on the other end of the connection. Since phone lines are full duplex connections, each modem can send and receive information simultaneously. The problem with these modems is that analogue telephone lines were never designed to carry digital information, modulated or otherwise, so voice-band modems tended to be very slow by today's data transmission standards, and made the telephone line unavailable for other communications.

Newer types of modems are now available to allow broadband communications over phone and cable television lines. Asynchronous Digital Subscriber Line (ADSL) modems use higher frequencies than the original voice-band modems, which allows for much higher transmission rates, as well as leaving the frequencies used for voice communication unused. This means that an ADSL modem can be used on a regular phone line without interrupting regular phone service. As the word "asynchronous" implies, ADSL modems are optimised for the downloading of data, at the expense of having reduced upload speeds. Cable modems make use of unused bandwidth available on the coaxial cables that provide television service to cable television subscribers.

As the Internet becomes more sophisticated, and costs come down, increasing numbers of users are choosing to have fibre-optic cables installed so that consumers and businesses are directly connected with the Internet.

Routers

A router extracts information on the destination of a packet it receives, selects the best path to the destination, and then forwards the data packets to the next device along the recorded path. They are used to connect separate areas (subnets) of the network, and control which information is routed to which subnet. Routers are used within and between enterprises and the Internet, and also within Internet Service Providers (ISP). The smallest routers provide connectivity for small networks, for example in small offices. The largest routers are used inside ISPs, or in the networks of large enterprises.

Routers generally contain a specialised operating system. High-end routers may include several processors and specialised application-specific integrated circuits (ASIC) and can undertake parallel processing. Chassis-based systems like the Nortel MERS-8600 or ERS-8600 routing switch also have multiple ASICs on every module and can work with a variety of LAN, MAN and WAN port technologies. Here, the chassis refers to the mounted framework holding the motherboard, memory, disk drives and associated equipment.

Firewalls

Firewalls are hardware or software devices configured to allow or exclude data from a network and prevent network intrusion to the private network. Firewalls regulate the traffic between computer networks, each of which may have different trust levels. For example, the Internet is a zone with no trust, while an internal network is a zone of higher trust, and a zone between these, with an intermediate trust level, might be a "perimeter network" or Demilitarised Zone (DMZ). Firewalls may use a "default-deny" firewall, where only explicitly allowed network connections can pass. Due to difficulties in defining these, the more common type is a "default-allow" firewall, where all traffic that is not explicitly blocked is allowed. This latter type however may allow inadvertent network connections. Different types of firewalls include:

- Network layer and packet filters, which prevent packets passing unless they match the filter rule set.
- Application layer firewalls intercept all Web browser, Telnet or FTP traffic, and may intercept all packets travelling to or from an application. Telnet is a network protocol used on the Internet to allow users to login to a remote terminal using a text interface; FTP refers to the File Transfer Protocol, which is used to transfer data from one computer to another over the Internet, or through a network.
- Proxy servers or proxy devices may function as firewalls depending on how they process or block particular types of input packets

Hubs, Switches and Bridges

A hub is a device for connecting Ethernet devices together, enabling them making them act as a single segment. Switches are similar to hubs, in connecting network segments, but are “smarter" in being able to inspect received data packets, and identify the source, destination device and forwarding requirements of the packet. Bridges are similar to hubs, but allow the management of the network traffic from one network to another.

3.4.3 Network Software

Network software allows computers to communicate with each other using network hardware. The software makes use of the hardware to create a path on which data can be transferred from computer to computer.

Transmission Control Protocol/Internet Protocol (TCP/IP)

Most computers use the TCP/IP suite of software for networking. TCP/IP stands for Transmission Control Protocol/Internet Protocol. TCP/IP translates the individual bits received by the networking hardware into highly organised and abstracted data that can be used by the Internet computer applications that we are familiar with, such as e-mail readers and Web browsers.

The TCP/IP model consists of a "stack" of four layers that progressively transform the raw bits from the network card into a more usable form. The lowest level of the stack is the host-to-network layer, which controls the transmission of data via a modem. The next layer is the Internet layer, which supports the Internet Protocol, which assembles the individual bits into a series of packets. A packet is a formatted block of data carried by a computer network, which is smaller than any individual file. The third layer is the Transport layer, which controls the transmission and reception of packets. Finally, the Application layer assembles the packets into an appropriate format, which can be read by computer applications.

In contrast with the TCP/IP model, the Open Systems Interconnection (OSI) reference model performs the same tasks of assembling individual bits into formats that can be read by computer applications, however, the OSI model uses seven layers instead of four. Whereas the TCP/IP model is based on the practicalities of day-to-day information transmission, the OSI model is more rigorous in its definitions of what each layer does. The TCP/IP model combines the physical and data link layers of the OSI model into the single most-to-network layer, and does away with the Session and Presentation layers of the OSI model entirely (Figure 14). Although the OSI model is better defined, it is used most often to explain the concepts behind network stacks and layers. TCP/IP is used as the real-world implementation of the concepts that are used in the OSI model.

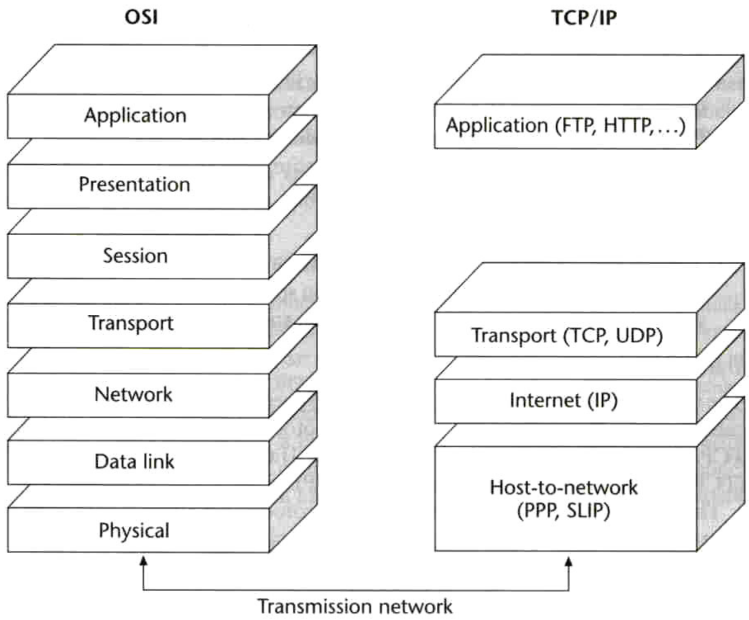

Figure 14. The OSI model as compared with the TCP/IP model (Source: Groot and McLaughlin, p. 125)

The Host-to-network layer in the TCP/IP model controls the transmission of data through a modem, and affects things such as data flow between the source and destination, the speed of transmission and rudimentary error checking to ensure that all bits have been received. The Point to Point Protocol (PPP) and the Serial Line Internet Protocol (SLIP) are standard protocols that are used to transfer the data.

Above this sits the Internet Protocol (IP) Layer, within which, the bits provided by the Host-to-network layer are assembled into packets. IP may be used over a heterogeneous or a homogeneous network. Heterogeneous networks connect computers with a mix of Ethernet and other types of connections. Each IP packet consists of 32 bits of raw data with 192 bits of additional information added to it (Figure 15). The additional information is used to control the behaviour of the packet, for example:

- The type of service. Different types of service have different specification, for example emphasizing speed or reliability
- The order of the packet, so that the data can be reassembled upon receipt
- The time-to-live parameter ensures that if the packet cannot be delivered, that it will eventually be destroyed, rather than circulating through the Internet forever
- The header checksum is used to determine whether any bits in the packet have been lost
- The source and destination address control the path that the data takes through the Internet

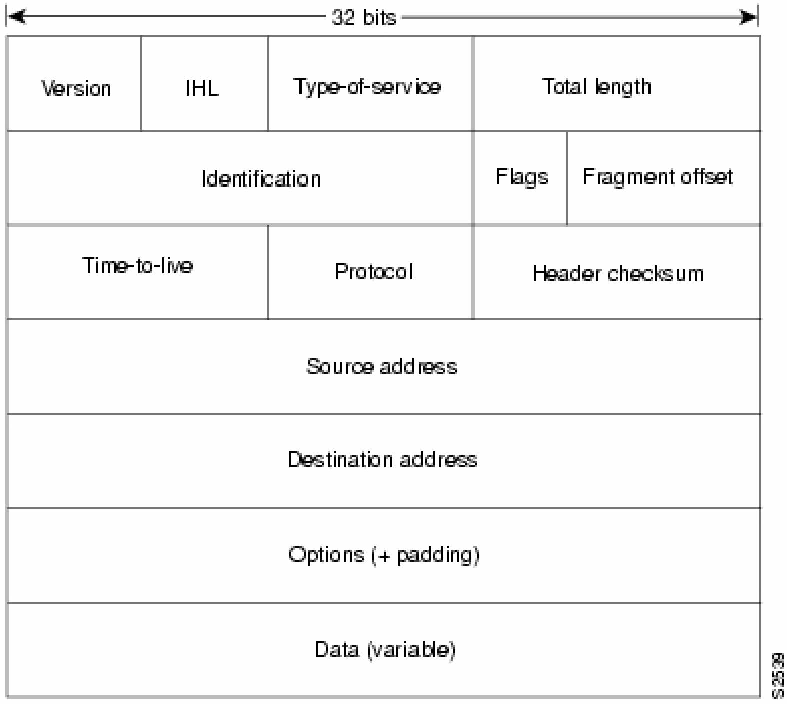

Figure 15. IP packet structure. Source: http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/ip.htm#wp2468

The Transport Layer controls how packets are sent. Whereas the Internet layer blindly transmits and receives packets, the Transport layer checks what is sent and received to make sure that everything is in order. There are two basic ways that the Transport Layer can operate: it can be optimised for speed, or it can be optimised for accuracy. In this layer, Transmission Control Protocol (TCP) or User Datagram Protocol (UDP) packets may be sent as a replacement for IP packets[2].

Transmission Control Protocol (TCP) is an accuracy-oriented protocol. It is much slower, but is designed to ensure that not a single bit of data is lost or changed from source to destination. For data files and programs, this protocol ensures that the files are not corrupted when they arrive. The Checksum value is used to determine whether any data has been corrupted during transmission. If the Checksum value does not match the data that is found in the packet, the Transport layer software requests that the packet be resent. For each 32 bits of information sent in a TCP packet, 192 bits of header information are required.

The User Datagram Protocol (UDP) is a speed-oriented protocol for the Transport Layer that is used when information needs to be delivered rapidly, but not necessarily accurately. UDP is used for Internet-delivered media such as streaming video and music, where errors in the transmission are acceptable. Unlike the TCP protocol, the UDP protocol simply assembles packets in the order that they arrive. The trade-off in not including the packet order is that the UDP Protocol has only 64 bits of overhead, which allows for much higher overall transmission speeds than TCP.

[2] If you find that the differences between the Internet and Transport layers in the TCP/IP model are unclear, then you can begin to understand why the OSI model explains the stack concept better. In TCP/IP, the only difference between the Internet and Transport layers is a result of the packet types used; the OSI model is much more logical in its separation of Data Link, Network, and Transport layers.

Much of what firewall software does occurs at the Transport layer. The Transport layer implements the concept of ports. Each TCP or UDP packet has a port number assigned to it by the originating computer. Firewall software typically shuts down a large number of these ports (there are hundreds) so that only particular types of packets are accepted. In addition, the software examines the packets that arrive at the ports that are left open, and may reject them if they are of the wrong type, come from the wrong address, or have other features that may make them “dangerous.”

The final layer in the TCP/IP stack is the Application layer. At the Application layer, the packets coming from particular ports are assembled together. The stream of data that results is of a particular type, and can only be used by a particular type of program. The data might be assembled into files and computer programs using the File Transfer Protocol (FTP), or it might be assembled into an e-mail message using Post Office Protocol 3 (POP3). There are many different protocols at the Application layer level. Some of the protocols that have been defined include:

- File Transfer Protocol (FTP) -- for transfer of files and programs
- Post Office Protocol 3 (POP3) -- for transfer of e-mail content
- Simple Mail Transfer Protocol (SMTP) -- for transfer of requests for e-mail
- Network News Transfer Protocol (NNTP) -- for transfer of the information from newsgroups
- Hypertext Transfer Protocol (HTTP) -- for transfer of Hypertext Markup Language (HTML) files from Web servers to Web browsers
- Domain Name Service (DNS) -- for transfer of translations between domain names, such as “google.com” and Internet Protocol addresses, such as 192.168.24.1.

The many different protocols at the Application layer level allow the transmission of virtually any type of data over the Internet, from movies to telemetry to telephone calls.

3.4.4 Types of Networks

Local Area Networks

Local area networks (LANs) are computer networks covering small geographic areas, like homes, offices, or clustered buildings. LANs provide high data transfer rates within small geographic ranges, and do not require leased telecommunication lines.

Smaller LANs generally consist of a one or more computers linked to each other -- often with one connected to a router and cable or ADSL modem for Internet access. Larger LANs may involve one or more servers, dozens of PCs, networked printers, and a dedicated fibre-optic link to the Internet. It is somewhat artificial to divide LANs by size today. Networks of different scales, or combined small-scale networks may be connected to each other to create larger networks. Ultimately, just about everything is now connected via the Internet.

An Ethernet is a frame-based computer networking technology used for local area networks, using physical wiring for connection, and hubs, switches, bridges, and routers to handle the transfer of data within the network. LANs use Ethernet cables for physical connections and IP as the network layer protocol.

While Ethernet is the most popular option for connecting computers in a LAN, there are a number of other options now available, however. Wireless Local Area Networks can be created, in which two or more computers are linked by radio. This has obvious advantages in buildings where it is difficult to route Ethernet cabling. Networks can also be routed through telephone systems using Home Phone line Networking Alliance (HomePNA) technology, whereby all the components of a computer network use existing telephone wiring in harmony with the voice and/or fax services. Power line communication is another option that uses existing copper wiring to carry information as well as power.

The Network File System (NFS) is a system for implementing a LAN over TCP/IP. NFS uses a different packet type than TCP, which is better adapted to its objectives of allowing the sharing of resources such files, printers and other resources.

Virtual Private Networks

Virtual private networks (VPNs) are communications networks that are tunnelled through other networks, but operate as a specific network. The difference between VPNs and other systems is the overlaying of other network(s) to provide a defined functionality useful to the community. An example is a secure communications system that tunnels through the public Internet. Some VPNs have security features, such as authentication or content encryption. VPNs without these features are used to separate the traffic of different user communities over an underlying network with strong security features. Different VPNs may have different service abilities. Some have a best-effort performance, which operates according to general specifications with no specific agreed methods. Others have a defined Service Level Agreement (SLA) between the VPN customer and the VPN service provider. The benefits of a VPN include:

- Lower costs than traditional routed networks over dedicated facilities, partly by outsourcing support and facilities
- Wider, faster links to different systems such as enterprise offices, as well as small-and-home-office and mobile workers
- Customisation of security and service operations for defined applications
- Scaling for variable demands of provider and customers in shared systems

However, problems include the requirement for more secure networks. The distribution of VPNs to multiple users (e.g. homes, telecommuters, and small offices) may allow access to sensitive information. Therefore, VPNs require effective security policies. This is especially the case when access is beyond the office facilities.

Wide Area Networks

Wide Area Networks (WANs) are networks that cover wide areas (metropolitan, regional, or national boundaries). The Internet is the largest and best known WAN. These may connect LANs, allowing communication between different local users. Some WANs may be private and belong to only one or more organisations. Some others are developed and maintained by Internet service providers, to allow connections between LANs and the Internet. Some WANs depend on leased telephone lines, and others depend on fibre-optic connections.

Leased lines have a router at one end, connecting to a LAN, and a connection to a hub within the WAN on the other end. Due to the high expense of leased lines, other methods such as circuit switching or packet switching methods may be used. A circuit switching network is one that establishes a fixed bandwidth circuit (or channel) between nodes and terminals before the users may communicate, as if the nodes were physically connected with an electrical circuit. A packet switching network is a network in which packets (discrete data blocks) may be routed between nodes, using links shared with other transmissions.

3.4.5 Functional Relationships

Client-Server architectures separate the client from the server, with each client or server connected to a network also defined as a node. The most basic has two types of nodes: clients and servers. The client software can send data requests to one or more connected servers, which in turn can accept these requests, process them, and return the requested information to the client. Most recent client software runs within web browsers.

The Client-Server architecture seeks stability by removal of complexity and possible uncontrolled changes in fundamental operations from underlying network components. This allows fixed numbers of clients and servers, with changes based on the administrator's decisions. This differs from active networking, which as seen below allows packets to modify the network.

The active network architecture is composed of a node operating system, connected to one or more executing environments, and active hardware (for routing or switching and executing codes within active packets). Executing environments are the systems that carry out the user requests. In contrast to traditional networks, active networks can change fundamental operations, using the network components. For example, the Atropos Toolkit allows experimentation with predictive capability within a network while the network is operating, changing bandwidth and memory size. An example of an active network is a peer-to-peer network.

Peer-to-peer systems use variable, ad hoc connections between network participants in a network, rather than conventional centralised resources where a few servers provide the service or application. Common applications include sharing content files containing audio, video, or digital data, and real-time data, such as telephony traffic. Pure peer-to-peer networks do not have clients or servers, but equal peer nodes functioning simultaneously as both clients and servers to other network nodes. This contrasts with a non-peer to peer transfer, such as an FTP server, where the client and server programs are distinct, with the clients initiating the download/uploads and the servers satisfying such requests. The key advantage of a Peer-to-Peer relationship is that the capacity of the network increases with the number of peers involved, since each peer computer contributes to the overall power of the network.

3.4.6 Network Topologies

Interconnections may be defined by network topology, the arrangement or mapping of the elements (links, nodes, etc.) of a network, especially the physical (real) and logical (virtual) interconnections between nodes. For example, any given node in the LAN will have one or more links to one or more other nodes in the network. The mapping of these links and nodes onto a graph gives the geometrical shape, or the physical topology of the network. These represent the physical connections, the wiring and cable layouts, node locations, and interconnections between nodes and cabling/wiring systems. The mapping of the flow of data between the network nodes determines the network's logical topology. Some common topologies (Figure 16) are listed below.

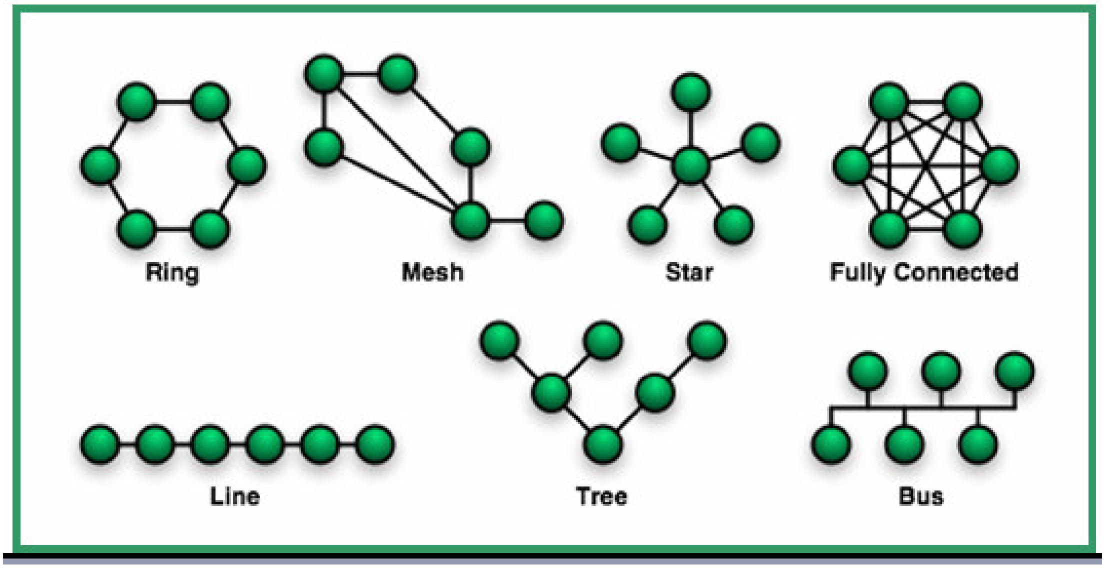

Figure 16. Network Topologies. Source: http://en.wikipedia.org/wiki/Image:NetworkTopologies.png

- Ring. In a Ring Network, each of the nodes of the network is connected to two other nodes in the network, with the first and last nodes connected, forming a ring shape. The transmitted data moves from one node to the next in a circular fashion. Usually this data flows in a single direction.
- Mesh. In a Mesh Network, the component parts can all connect to each other through multiple hops. The advantage of a meshed network is dependent on the number of components.
- Star. In a Star Network, each of the nodes of the network is connected to a central node in a hub and spoke formation, with central node as the hub and attached nodes the spokes. Data transmitted between nodes in the network goes to the central node, which may or may not retransmit the data to other nodes in the network.
- Fully connected. In this type of network, each of the nodes of the network is connected to each of the other nodes with a point-to-point link. This enables the simultaneous transmission of data from any node to all others.
- Tree (or hierarchical). Here the central or root node at the top level of the hierarchy, is connected to one or more other nodes one level lower in the hierarchy. There is a point-to-point link between each of the second level nodes and the top level central node. The second level nodes are connected to one or more nodes one level lower in the hierarchy (third level). The tree hierarchy is symmetrical, with each node having a fixed number of nodes connected to it at the next lower level.
- Line (Point-to-Point). This is the simplest topology, comprising a permanent link between two endpoints, this guaranteeing communication between the two endpoints.
- Bus. This is usually divided into two types, the linear bus and the distributed bus. In the linear bus all of the nodes of the network are connected to a common transmission medium with two endpoints (the bus, backbone, or trunk), and all data transmitted between nodes in the network is transmitted over this common transmission medium. The data can be received by all nodes in the network simultaneously. In the distributed bus, all of the nodes of the network are connected to a common transmission medium that has more than two

endpoints, created by adding branches to the main section of the transmission medium. Like the linear bus, all nodes share a common transmission medium.

These systems may be combined, this creating hybrid networks that combine the features of two or more typologies. The resulting network does not exhibit one of the standard topologies (e.g., bus, star, ring, etc.). Two common examples of Hybrid networks are star-ring networks and star-bus networks.

3.4.7 Security, Authentication, and Authorisation Services

Computer security is a branch of computer science concerned with the enforcement of security in computer operations. It also involves the protection of assets from access, change, copy, theft, or destruction. In computing there are usually three forms: physical security, virtual security and data security. The level of security varies according to applications and requirements. Generally, there is an implicit or explicit statement or policy defining confidentiality, integrity and availability of processed or stored electronic information on computer software, hardware and networks. Software is designed to make use of specifically defined hardware and software mechanisms that control the operating system within the security policy, in some cases restricting access and privileges of users.

- Physical security. This includes locks on the server or mainframe room, environmental control, the nodes of the wired network, the condition of the nodes (e.g. power and cooling) and protection of wiring from water and traffic. This prevents theft as well as hacking.
- Virtual security, either wired or wireless is concerned with the external world to which an organization is connected. Wired virtual security protects the organization's data from access through wired ports, such as the Internet. Access may take the form of hackers, viruses and spam. Wireless virtual security includes all of the wired virtual security, as well as access from wireless sources.
- Data security is the objective of both physical and virtual security.

Authentication is the process verifying the digital identity of a sender (termed the principal, either the person using a computer, a computer itself or a computer program) of a digital request of a communication such as a request to log in. This ensures that the principal is the user authorised to receive a request. This contrasts with a blind credential, which does not establish principal identity, but gives a defined status to the principal. Authentication contrasts with authorisation, as authentication is the process of verifying a person's identity, while authorisation is the process of verifying that a known person has the authority to perform the operation. Therefore, authorisation cannot occur without authentication.

Authorisation protects computer resources by allowing only authorised users to have access to the data and/or processes of the system. This authorisation process verifies that a digital communicator has been granted permission to use the system. Permissions may be defined by the administrator with an access control list or on the principle of least privilege, which grants minimal permission for the principal's job. Anonymous consumers or guests, who may be common in a distributed system, are usually not authenticated and have few or no permissions. Maintaining the security policy is necessary, and requires regular updating of software. It is often desirable to remove a user's authorisation: to do this with security policy application requires that the permissions be updateable.

Security and Encryption

Hypertext Transfer Protocol (HTTP)

This is a communications protocol used to transfer or convey information on intranets and the World Wide Web. Its original purpose was to provide a way to publish and retrieve hypertext pages. This is a request/response protocol between a client and a server. The client or user agent (e.g. a web browser) makes a HTTP request to the responding or origin server (which may store or create resources such as HTML files and images). Between the user agent and origin server may be intermediaries, including proxies, and gateways. HTTP typically uses TCP/IP and its supporting layers, but may also use any other protocol on the Internet, or on other networks. An HTTP client initiates a request by starting establishing a Transmission Control Protocol (TCP) connection to a particular port on a host. The HTTP server on that port waits for the client to send a request message, and when it receives the request, the server sends a status line, for example "HTTP/1.1 400 OK", and another file, which may be the requested file or an error message. The resources to be accessed by HTTP are identified using Uniform Resource Locators (URLs).

Transport Layer Security (TLS) and Secure Sockets Layer (SSL)

TLS and SSL protocols provide secure communications on the Internet for activities such as web browsing, e-mail, Internet faxing and instant messaging. The TLS and SSL protocols are broadly similar. The TLS protocol enables application communication across a network, designed to eliminate eavesdropping, tampering, and forged messages. It also provides authentication and communications privacy over the Internet. Usually, only the server is authenticated (i.e., its identity is determined) while the client remains unauthenticated. Thus, the end user (an individual or an application, such as a Web browser) is assured of the identity of the server. The next level of security is termed mutual authentication. Here both ends of contact are certain their contact's identity. TLS involves three basic phases:

- Peer negotiation for algorithm support
- Key exchange and authentication
- Symmetric cipher encryption and message authentication

The Secure Sockets Layer (SSL), developed by Netscape, was an earlier protocol for managing the security of a message transmission on the Internet. Transport Layer Security (TLS) has replaced SSL. SSL uses a program layer that is located between the Internet's Hypertext Transfer Protocol (HTTP) and Transport Control Protocol (TCP) layers, and is included as part of both the Microsoft Internet Explorer and the Mozilla Firefox browsers and most web server products. The "sockets" refers to the sockets technique of passing data back and forth between program layers on a computer, or between a client and a server program in a network.

Digital Identity Certificate

This is also known as a public key certificate, and is an electronic document that includes a digital signature to tie a public key with an identity (i.e. the name of a person or organization and their address). This certificate verifies the individual owner of a public key. In most Public Key Infrastructure (PKI), the signature is of a Certificate Authority (CA). In a web of trust scheme, the signature may be of a user (a self-signed certificate) or other users ("endorsements"). In both cases, the certificate signatures are indications that the identity information and the public key belong together.

Such certificates are convenient for large-scale public-key cryptography, as securely exchanging secret keys between users is impractical for large networks.

A certificate usually includes:

- The public key being signed
- A name, which can refer to a person, a computer or an organization
- A validity period
- The location (URL) of a revocation centre
- The digital signature of the certificate, produced by the CA's private key

Digital Signatures

These are used to simulate the security properties of a signature in digital, rather than written form. Digital signature schemes usually have two algorithms, one for signing which involves the user's secret or private key, and another for verifying the signatures of the user's public key. Like written signatures, digital signatures are used to provide authentication of the input or message. Digital signatures are used to create public key infrastructure schemes, where a user's public key is tied to a user by a digital identity certificate issued by a certificate authority. PKI schemes also link user information (name, address, phone number, etc.) to a public key, so that public keys can be used as a form of identification. A digital signature scheme usually consists of three algorithms:

- A key generation algorithm producing a "key pair" (PK, SK) for the signer, where PK is the verifying (public) key) and SK is the signing (private) key
- A signing algorithm (S), that on input of a signing key (SK) and a message (m) produces a signature (σ)
- A signature verifying algorithm (V), that on input a message (m), a verifying key (PK), and a signature (σ), either accepts or rejects

Digital signatures may also be used to implement electronic signatures. In some countries, including the United States, and in the European Union, electronic signatures are legally recognised.

Types of Portal Attacks and Defences

These concern internal or external threats to computer and network operation. Threats include users, programmers, hardware, databases, systems software, viruses and Trojan horse programs. The results of these threats may be theft of data, destruction of resources, disruption of service, fraud, loss of business reputation, stoppage of work, and/or equipment sabotage. Attacks may also include denial of service and distributed denial of service. Defences discussed below include hardware and software systems.

Denial of Service Attack

A Denial-of-service (DoS) attack is an attempt to make a computer resource unavailable to its intended users. These are usually the planned and malevolent efforts of people to prevent an Internet site or service from functioning, temporarily or indefinitely. Such attacks may target sites or services linked to high-profile web servers such as banks, credit card payment gateways and DNS root servers. DoS attacks are usually implemented by:

- Forcing the targeted computer(s) to reset, or consume its resources such that it can no longer provide its intended service; and/or,
- Obstructing the communication media between the intended users and the target so that they can no longer communicate adequately

They may take several forms, including:

- flooding a network, thereby preventing legitimate network traffic;
- disrupting a server by sending more requests than it can possibly handle, thereby preventing access to a service;
- preventing a particular individual from accessing a service;
- disrupting service to a specific system or person;
- consumption of computational resources, such as bandwidth, disk space, or CPU time;
- disruption of configuration information, such as routing information;
- disruption of state information, such as unsolicited resetting of TCP sessions;
- disruption of physical network components; or
- obstructing the communication media between the intended users and the victim so that they can no longer communicate adequately

DoS attacks are violations of the Internet proper use policy and may violate of the laws of individual countries.

Distributed Denial of Service Attacks (DDoS).

These occur when multiple compromised systems flood the bandwidth or resources of a targeted system, usually one or more web servers. Attackers using a variety of methods compromise these systems. The major advantages to an attacker of using a distributed denial-of-service attack are that multiple machines can generate more attack traffic than one machine, multiple attack machines are harder to turn off than one attack machine, and that the behaviour of each attack machine can be stealthier, making it harder to track down and shut down. These attacker advantages cause challenges for defence mechanisms. For example, merely purchasing more incoming bandwidth than the current volume of the attack might not help, because the attacker might simply be able to add more attack machines.

Firewalls

These are hardware or software devices configured allow or prevent data through a computer network. Firewalls may restrict the spread of networked computer worms and Trojan Horses (Trojans). Variations in firewalls depend on the location of the communication process and its interception.

- Network layer firewalls, also called packet filters, prevent passage of packets that do not match the established rule set (either defined by the firewall administrator or by default rules). Network layer firewalls are usually of two types: stateful and stateless.
  - Stateful firewalls maintain information, termed ‘state information’ about the source and destination IP address, UDP or TCP ports, and attributes of the connection such as session initiation, data transfer and/or completion connection). Definition of this information and the mapping allows an assessment of the packet. If a packet does

not match an existing connection, it will be evaluated according to the rule set for new connections. If however, a packet matches an existing connection based on comparison with the firewall's state table it is allowed to pass through.

  - Stateless firewalls have packet-filtering capabilities, but cannot make decisions that are more complex on the packages.

- Application layer firewalls work on the application level, and may intercept all packets travelling to or from an application, and block other packets.
- Proxy servers may act as a firewall by evaluating input packets (e.g. connection requests) while blocking other packets.

Firewalls often have network address translation (NAT) functionality to hide the true address of protected hosts. The hiding the addresses of protected devices has become an increasingly important defence against network reconnaissance.

Honeypots

These are traps set to detect and counter attempts at unauthorised use of information systems. Such systems consist of a computer, data or a network site, set up to look like part of a network but are isolated and unprotected, and may be configured to be valuable to online attackers. Any information attracted may then be surmised as unauthorised. The honeypot is valuable as a surveillance and warning tool for viruses, spam and other online attacks. In some cases if honeypots are not properly walled off and monitored, an attacker can use them to break into a system.

Perimeter Networks

Perimeter Networks, also known as Demilitarised zones (DMZs) or demarcation zones. These are physical or logical sub-networks including the organisation's external links to a larger, less trusted network, such as the Internet. This gives an additional layer of security for the LAN. The most vulnerable hosts to attack are those used by users outside the LAN (e.g. e-mail, web and DNS servers). For protection of hosts and the network against a successful invader, these hosts are placed into their own sub-network.

User Authentication Procedures

Passwords

These are secret (from unauthorised users) authentication data used to control access to a resource. Common methods of password protection include:

- Obscuring the password as it is typed by using asterisks or bullet characters
- Forcing users to create passwords of minimum length e.g. Unix systems may limit passwords to eight or more characters
- Requiring users to re-enter their password after an inactive period
- Enforcing a password policy requiring difficult to break (strong) passwords
- Requiring periodic password changes
- Assigning random passwords
- Using encrypted password-authenticated agreement to prevent network attacks on transmitted passwords

Forced password changes

This requires users to change passwords periodically, e.g. every 3 or 6 months. Such systems may prevent users from choosing a password similar to a previous choice. This however, may encourage choice of weaker passwords when the users run out of possibilities. An alternative is a very strong password, changed infrequently, but this may cause problems if an unauthorised user learns it. Sanctions for disobeying password policies may progressively include warnings, loss of computer privileges or job termination. If confidentiality is legally mandated, e.g. when dealing with classified information, violation of password policy can lead to criminal prosecution.

Ensuring secure passwords

Strong passwords are sufficiently long, complex, randomly chosen and/or producible only by its creator. These attributes should ensure that unauthorised acquisition would be too difficult to be worthwhile. The difficulty will vary with the attacker and his/her resources and the value of access to the attacker. For example, learning a password controlling bank access may be worth months of computer time, but that of a child's computer may not be worth the effort. Even strong passwords may still be stolen, tricked or extorted from an owner, intercepted in transit, or even accidentally discovered by another user.

Biometrics

This involves the study of methods of individual human recognition, related to one or more physical or behavioural identifiers. This may be possession-based (i.e. using one specific item such as a card or security tag) or knowledge-based (i.e. the use of a code or password). Standard validation systems may use multiple inputs of samples for sufficient validation. Attributes considered in validation may be based on:

- Uniqueness, based on individual recognition
- Permanence, resistance to aging
- Collectability, the ease of acquisition for measurement
- Performance accuracy, speed, and utility of the technology
- Acceptability or the degree of approval of a technology
- Circumvention, the ease of use of a substitute

Security Analysis Tools

These generally collect information about networked hosts, enabling system administrators to use this information to test their systems for vulnerability to unauthorised attacks from the network. However, they may be useful for both good and bad uses, as intruders may also determine the weakness of s system with their processes. There are several security analysis tools. For example, Nessus is the world's most popular vulnerability scanner, estimated to be used by over 70,000 organisations worldwide. It was first in the 2000, 2003, and 2006 security tools survey (by SecTools.Org). It is a comprehensive vulnerability-scanning program, detects potential or confirmed weaknesses on tested machines. On Unix systems (including Mac OS X), Nessus runs as a daemon (nessusd). A daemon is a computer program that runs in the background, which has a process number 1 (init short for 'initialization' that spawns all other processes) as its parent process and no controlling terminal). The Nessus daemon scans the systems and Nessus (the client) controls scans and communicates the vulnerability results to the user. Unix systems, as operating systems for medium and large computers, are composed of several components, including the development environment, libraries, documents, and portable, modifiable source-codes for all components). For Windows, Nessus 3 has a self-contained scanning, reporting and management system. Problems tested by Nessus include:

- Gaps in security allowing a remote cracker to control the machine or access sensitive data
- Denial of service to unauthorised users
- Misconfiguration (e.g. open mail relays)
- Testing of unapplied security patches, even where any flaws are unexploited in tests

Nessus may also engage an external tool called Hydra to launch a dictionary attack against a system that is being tested. A dictionary attack is a method for passing a cipher or authentication mechanism by determining its decryption key or passphrase. It searches a very large number of possibilities, but differs from a brute force attack (which searches all possibilities) as it tries only possibilities that are very likely to succeed (e.g. derived from a list of words in a dictionary, because many people choose short simple passwords of less than eight characters).

When operating, Nessus does a port scan with one of its four own internal port scanners to determine open ports on the target and attempts to exploit these open ports. The resultant vulnerability tests, available as subscriptions, are written in NASL (Nessus Attack Scripting Language).

Another security analysis tool is SATAN (the Security Administrator Tool for Analysing Networks), developed in 1993. This was the first truly user-friendly network scanner, featuring an HTML interface, with forms to enter targets, tables to display results and context-sensitive tutorials that appear on the discovery of a hole. It is especially useful for multiple hosted, networked systems. SATAN uses a web browser (e.g. Internet Explorer or Firefox) to provide the user interface. In addition to vulnerability reporting, it also collects general network information. This may include information on the hosts connected to subnets, the machine types and the services offered.

3.4.8 Servers and Network Balancing

Proxy Server

Proxy servers forward client requests to other servers. The client connects to the proxy server, requesting a service (e.g. a file, connection or web page) that is available from a different server, and the proxy server connects to the new server and requests the service for the client. A proxy server may also change the client's request or the server's response. It may also serve the request without contacting the new server, by caching the first request to the remote server, saving the information for later, and speeding up the process. A proxy server that passes all requests and replies unmodified is called a gateway or tunnelling proxy. Proxy servers may be located in the user's computer or at specific points between the user and the other servers or the Internet.

Reverse Proxy Server

These are proxy servers linked to one or more servers, and usually in front of web servers. All connections from the Internet to a web server pass through the proxy server. The proxy server may deal with the request itself or pass the request to the main web servers. The reverse proxy server presents a single interface to the caller. This contrasts with a forward proxy that acts as a proxy for out-bound traffic.

Reverse proxy servers are installed mainly for the following reasons:

- Security: an additional layer of defence protecting the web servers
- Encryption: this may be done by a reverse proxy that is equipped with acceleration hardware
- Load distribution: the reverse proxy can distribute the load to several servers
- Caching static content: offloading the web servers by caching static content, such as images
- Compression: the proxy server can optimise and compress the content to speed up the load time

Network Load Balancing

Network Load Balancing (also called dual-WAN routing or multi-homing) balances traffic across two WAN links without using complex routing protocols. This balances network sessions like web, email, etc. over multiple connections to spread out the amount of bandwidth used by each LAN users, increasing the total amount of available bandwidth. Network load balancing is also used to provide network redundancy, as during a WAN link outage, access to network resources will still be available through the secondary links. Most network load balancing systems also include the ability to balance outbound and inbound traffic. This may reduce costs and ensure effective control.

3.4.9 Lessons Learned

In this module, we have learned about the different characteristics of computer networks. Computer networks are very complex. While all the different systems described strive for easy use, speed, accuracy, low cost and reliability, many different methods and networks have been developed. Computer networks must balance several important issues:

- Access to networks for the maximum number of authorised or free users, for the benefit of society
- Restriction of access to sensitive or private information, and the reliability of such regulative mechanisms
- Connectivity between different components of networks, with problems based on size, scale, complexity, use and hardware availability
- Increased usage of computer networks, which requires greater sophistication and usability

Computer networks are constantly developing, in reaction to changing environments. Therefore, Systems Administrators, Network Designers, and Computer Security Experts must constantly be improving their skills to keep up with changes in technology.

3.5 GII Software Components

In this and the next lecture, we will discuss the computer software that makes GII possible. There are many different levels of software that are interconnected in a GII. Software may be "commercial off the shelf (COTS)" or it may be custom-built to support GII functionality. In this lecture, we will examine COTS software products, and in Module 4, Lecture 1, we will examine how custom programming will be used to combine all of the components of the Lithuanian GII to create a single system of hardware, networks, and software.

Virtually any kind of server software can be combined in the GII, including software for web page display, database management, creation and maintenance of metadata, performing financial transactions, and for performing GIS functions such as data format conversion, map rendering and satellite image processing. All of these elements are united into a single portal, which provides internal and external users with a powerful set of tools to effectively utilise the hardware and software capabilities of the GII (Figure 17). Portals make use of web server software, which presents web pages to online users.

Figure 17. Canadian Geospatial Data Infrastructure GeoConnections Discovery Portal (http://geodiscover.cgdi.ca)

3.5.1 Web Server Software

Every time that you access a web site on the Internet, your browser is sending out a number of requests for information to a web server. Each web site consists of many different web pages, plus associated graphics and other files. A web server is, in its purest form, a server that retrieves pages of information and makes them available to a browser over the Internet. This is a fairly specialised task which requires web server software to handle the requests effectively. Increasingly sophisticated web sites and have required more powerful Web Servers, so modern web servers provide not only pages of information, but also associated multimedia content, including image, sound, music and video files.

Two types of web server software currently dominate the market. The Apache Web Server and Microsoft IIS (Internet Information Server) together control 85% of the web server software market (Wikipedia article “Web Server”). Whereas IIS runs only on Windows platforms, Apache runs on Unix, Linux or Windows operating systems.

Operating a web server is relatively straightforward. Once the web server is configured (which used to be a difficult task, but has now been made much easier in later versions of Apache and IIS), the server administrator needs to adjust the operation of the web server software to optimise performance. Another administrative task is to ensure that the web server remains secure, which requires regular reviews of the security policies and the monitoring of security logs to ensure that the web server's security has not been compromised (Ince, 2004).

How Web Servers Work

Web Server software, as well as many other programs that form the Internet, make use of the Transport Layer (Layer 4) that is defined in the TCP/IP and OSI stacks (see previous lecture). The transport layer sends and receives information to and from the Internet in the form of packets. Different kinds of packets are used to transport different types of information, and each packet type is associated with a particular Transport Communication Protocol (TCP)/User Datagram Protocol (UDP) port on a computer. The ports can be opened and closed by firewall software to control the types of information that is received by the computer (see http://en.wikipedia.org/wiki/List_of_TCP_and_UDP_port_numbers#Ports_0_to_1023). Table 3 shows some of the most commonly used ports.

Table 3. Some commonly used TCP/UDP ports and the types of packets that they receive.

Port	Protocol	Use
80	Hypertext Transport Protocol (HTTP)	Sending web pages
20	File Transport Protocol (FTP)	Sending file data
21	File Transport Protocol (FTP)	Controlling transmission of file data
25	Simple Mail Transport Protocol (SMTP)	Sending e-mail messages between servers
110	Post Office Protocol 3 (POP3)	Receiving e-mail messages from mail server
443	Hypertext Transport Protocol (HTTP) over Transport Level Security (TLS)/Secure Sockets Layer (SSL)	Encrypted web pages
1723	Microsoft Point to Point Tunnelling Protocol (PPTP)	Virtual Private Network (VPN) connections

Web Servers work by sending and receiving from port 80 of a computer. When a user requests a document in a web browser, they identify the protocol, computer name, and document name. For example, the following request can be broken down as follows:

http://photos.igougo.com/pictures-I216-Lithuania_photos.html

Protocol | Computer Name | Document Name

First, the computer is identified. In the above example, the computer "photos" on the domain igougo.com is accessed. A request is sent to this computer for an http document. Because this request begins with http://, the network software directs this request to Port 80, where the web server software interprets the request. The web page "pictures-I216-Lithuania_photos.html" is retrieved by the web server software and is returned to the web browser from which the request originated (Figure 18).

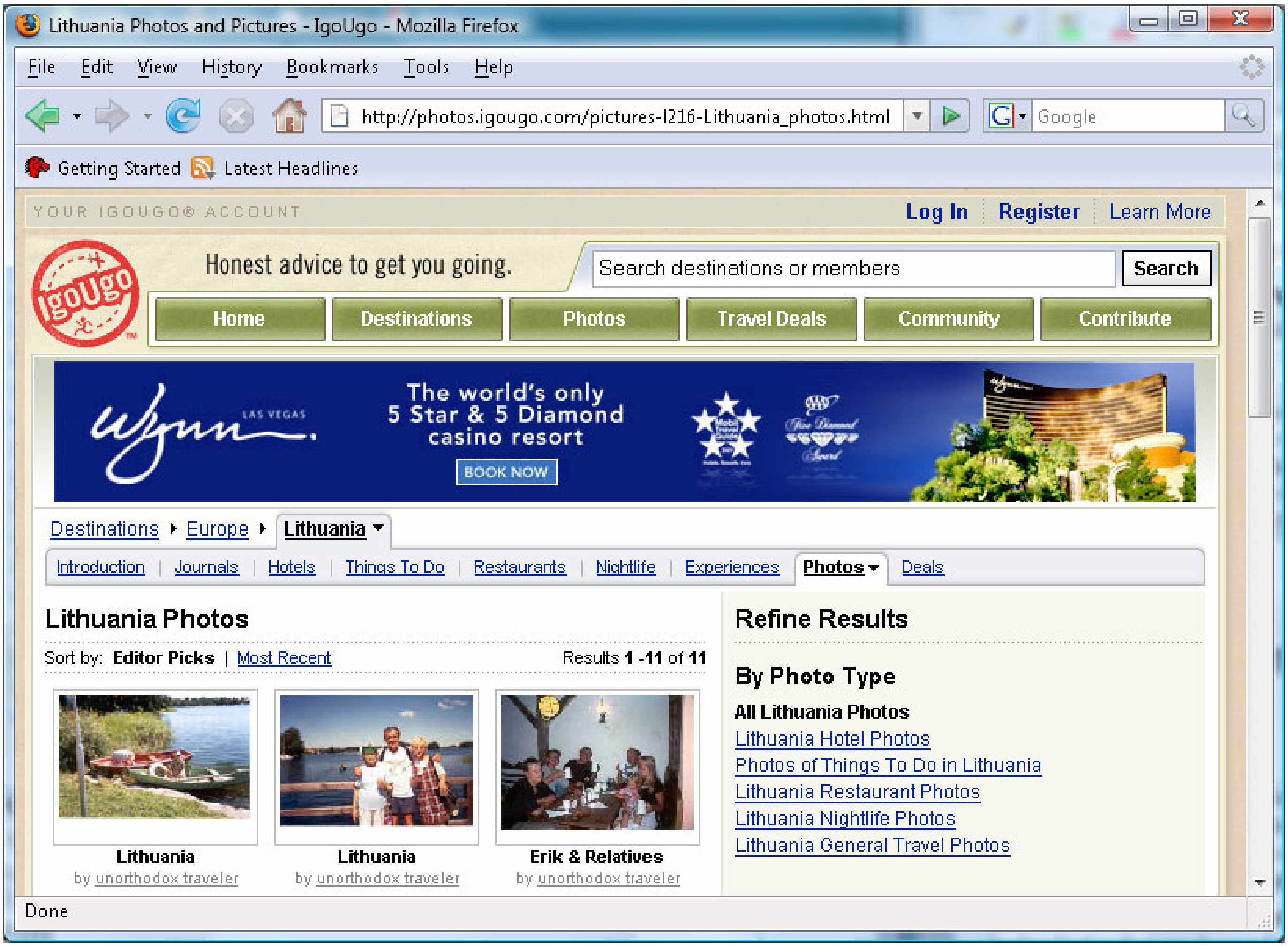

Figure 18. The web page "pictures-I216-Lithuania_photos.html" from the computer "photos" on the domain "igougo.com."

Not only HTML, but other files using protocols such as Extensible Markup Language (XML) may also be distributed by web servers. Web servers may also provide encrypted web pages on port 443 using the HTTPS (Hypertext Transfer Protocol – Secure) protocol, which encodes the HTML pages using Secure Sockets Layer (SSL) or Transport Level Security (TLS).

Dynamic Web Pages

Originally, web pages were static, and presented information that never changed. The need to have dynamic web pages has given rise to a number of technologies that can be used to change a web page on the basis of the content of a database. This allows, for example, a web page to display updated weather reports as soon as information from automated weather stations is stored in a database.

Scripts (simple programs) can be run within the user's web browser (client-side scripting), or on the web server itself (server-side scripting) to manage the updating of the web pages. Client-side scripting can be accomplished with Java applets or scripting languages such as JavaScript, Jscript, or VBScript. Active Server Pages or ActiveX are other client-side options. Server-side scripting is accomplished with techniques such as Java Server Pages (JSP) or languages such as PHP (a recursive acronym for PHP Hypertext Processor).

Web server software is one of the core technologies that allow a GII to exist. Web servers may be incorporated into Web Portal software (as in IBM WebSphere), or they may operate separately (Microsoft SharePoint/Internet Information Server). Web servers are the applications that web browsers directly interact with, and these serve to present information from other types of software that are hidden from the user. Chief among the hidden software are Database Management Systems.

3.5.2 Database Management Server Software

The Database Management Systems (DBMS) used on servers are closely related to desktop Database Management Systems. As you might guess, the DBMS software used shares many of the characteristics of other server-based software and hardware, and is designed using the same principles of fault tolerance and error recovery. Examples of database management server software include SQL Server, by Microsoft and mySQL, which is a popular Open Source server database (Ince, 2004).

As with many database management systems, database management servers provide the ability to produce numeric reports and display them online. This capability allows the GII to produce not only cartographic output, but also numerical reports created from the data in the GII database.

Structured Query Language

Like its PC-based brothers, server-based DBMS software uses relational database technology, and makes use of the Structured Query Language (SQL or "Sequel") to add, modify, and delete data. Unlike PC-based DBMS however, SQL is the sole interface to server-based DBMS; no supplemental Graphical User Interface (GUI) is provided to simplify the operation of the DBMS. Instead, SQL acts as the interface between the web pages and the DBMS. Using client or server-based scripting, the web pages are used to create SQL statements that are used by the DBMS for processing. The results of these transactions are then formatted and presented to the user by the same scripts that created the SQL statement.

SQL syntax takes the form <Statement> <Table> <Condition>, as can be seen in the following Select statement:

Select * from subscribers where amount_billed > 5000;

Columns to Select | Table to Select From | Condition to Meet

In this example, we are using the Select statement to display part of a table from a database. The wildcard symbol (*) indicates that all columns are to be displayed; alternately we could specify a list of column names separated by commas. “Subscribers” is the database table from which we are selecting. This is one of many different tables that might exist in a database. The where clause indicates which rows are to be retrieved; in the example only those rows will be retrieved in which the column amount_billed has a value greater than 5000.

By altering the rows selected and the where clause, any combination of rows and columns in the table can be displayed. Part of the power of SQL stems from its simplicity and its similarity to spoken languages. Only eight SQL statements are required to perform virtually all data manipulations; these are shown in Table 4.

Table 4. Basic SQL statements

Statement	Function
Select	Display a combination of rows and columns from the database that meet particular criteria
Insert	Add rows to a database table
Delete	Remove rows from a database table that meet particular criteria
Update	Change currently selected values
Commit	Make changes to the database permanent
Rollback	Undo changes to a database that have not yet been committed
Grant	Grant permissions to a particular user or group of users
Revoke	Revoke permissions from a particular user or group of users

Commit and Rollback in Distributed Systems

The Commit and Rollback statements deserve special mention in the context of web-based server systems. Many database management systems allow statements to be committed automatically, using the “autocommit on” statement. This works well in an environment with a few users entering SQL statements interactively. The risk of automatically committing statements is outweighed by the risk of having a user forget to issue a commit statement before exiting the DBMS software and losing all changes that have been made. In a web-based system, however, the commit and rollback statements take on special meaning. Since all of the manipulations are being orchestrated by computer programs, there is no risk of a commit statement being missed, unless a computer program crashes. In such an environment, we can make use of the commit and rollback statement to enhance reliability. Before issuing a commit, the controlling program can perform checks to ensure that the database is in the correct state. If a new user was added to a database, for example, a select statement could be run to ensure that the new user has been recorded in the database before a commit statement is issued. If this check fails, then a rollback statement can be issued to undo all changes, and the SQL commands to add the new user can be run again.

This might seem overly elaborate for a simple database transaction, but consider that the controlling program may be making hundreds of different changes in different databases. Every transaction must be entered correctly, or it will be very difficult to figure out later which statements worked and which failed.

Stored Procedures

Many transactions in a web-based system such as the updating of metadata in the Lithuanian GII will be very long and will involve many tables. Current database management server software can manage complex SQL statements as stored procedures. A stored procedure is a complex, commonly used SQL procedure which is used on a regular basis. Rather than recreating and transmitting such lengthy commands to the database server, and then having them optimised and executed, the optimised code can be stored directly on the database server ready to execute. This is analogous to a subroutine in a programming language – only one or two parameters need to be changed each time the stored procedure is run, and these can be provided to the stored procedure as variables when it is run. Using stored procedures reduces network traffic and allows optimised code to be run immediately, reducing the load on the database server (Ince, 2004).

Database Server Characteristics

As we described above, database servers are designed to be robust. Accordingly, many of the characteristics of database servers stem from these requirements. Obviously, a database server is designed to accept and process SQL queries. However, there are many different ways to execute SQL queries, some of which take much time, and others which are very quick. A database server will parse the SQL queries that have been entered, and will rewrite them in order to execute them most efficiently. The database server can do this because it has an internal model of the data structures that are stored on it.

Database server concurrency functions support multiuser database use. For example, many different requests from different users may arrive at the same time, so the database server needs to be able to handle these requests concurrently. In most cases, the SQL statements issued are simply Select statements, so reading the contents of the database for separate users at the same time is not difficult, however, when one or more users are updating the database, the issue becomes much more complex.

Potentially, a user may query data from a table at the same time that another user is making updates to the table. If the processes were executed by the database server (exactly) concurrently, then the user requesting data might receive half of the data requested based on the old data, and half based on the updated data. The results of the query in such a case would be incorrect. In such a case, the database server should lock access to the database until the updated information has been entered. Because of the nature of SQL, the data entry is separated from the actual insertion into the database, which in most cases only takes a fraction of a second. Thus, the user requesting the data would have his or her request delayed until the update has been completed, and since the database server operates so quickly, the delay would be virtually unnoticeable.

The situation becomes even more difficult if two users are updating two tables at the same time. In such a case, one process may lock the first table for editing and then request access to the second table; the second process may lock the second table and then request access to the first. If the timing of these requests is just right, then we end up with a deadlock, in which neither process is able to complete, and each ends up waiting for the other forever. To resolve such a condition, the database server software needs to rollback both of the transactions, execute one and allow it to complete, and then execute the other once the first has completed.

Database servers are also configured to ensure data security. This may be less of an issue when the database server is being driven by web-based applications, since all of the database requests are coming from another computer program, which may handle security on its own. In such a case, database security would be configured to accept requests only from the controlling computer program and from the database administrator. This helps to prevent outside users from gaining direct access to the database, in case the link between the controlling program and the database server is compromised.

Finally, database servers must be able to support backup and recovery procedures. While the transaction management capabilities afforded by the SQL commit and rollback statements are able to prevent most sources of database corruption, there are always unanticipated problems such as hardware failures that can corrupt a database, and so database administrators should be prepared to restore the database to an uncorrupted state in the event of a disaster (Ince, 2004).

Distributed Databases

Although SQL statements may be used to manipulate a number of different databases, it is also possible that there will be a single distributed database running the GII system. In a distributed database, a single SQL interface manipulates tables of data stored on different computers.

There are two reasons why we might want to construct a distributed database. The first has to do with performance. If your organization has groups working at a number of different locations, transmitting data from a central computer to the different locations may cause poor performance. The alternative is to have the tables housed in the same building as the group that uses them. This ensures that the tables required by each group can be rapidly updated, but that all of the data can still be accessed from a central computer. The second reason is simply because the world is becoming increasingly interconnected. Organizations and groups that used to operate by themselves are now working together, and smaller companies are being bought out by larger ones, so individual standalone databases are now often combined into distributed databases.

When tables or portions of tables are located in many different physical locations, the software that drives the distributed database can be quite complex. It must be able to manage and reconcile replicated and changed data in all the locations. Obviously, the job of managing distributed databases such as these is much more difficult than when all the tables reside on a single computer.

There are three ways of distributing the information between computers. The simplest is downloading, in which copies of database tables are made on a regular basis, and updates are synchronised back to the source of the data on a periodic basis. This solution only works when the data in the database are not time critical. The second method is replication, in which each change is copied to a number of locations automatically, as soon as the change is made. Replication works better for time-sensitive data, since each copy of the data is always kept up to date. The third way of distributing information is fragmentation, in which portions of tables are kept in different locations. Fragmentation may be horizontal, meaning that tables are divided by rows, and different rows are kept in different locations, or vertical, meaning that different columns are kept in different locations. Horizontal fragmentation might be appropriate in an organization with regional offices, in which each office performs the same basic task for its individual region. Vertical fragmentation is more appropriate when the database is split between organizations with different functions. Each office is provided with the columns that its employees need to perform their duties.

For a distributed database to be effective, it must offer the same rapid performance that users have come to expect from standalone databases. This means that an SQL query must be able to retrieve data from different locations as if it were coming from a single computer. The database must be able to avoid corrupting replicated data because of time lags or transmission failures. This is a fairly difficult task, given the number of events that must execute flawlessly in the correct sequence; for this reason the clocks on all of the computers that are involved in the distributed database must be synchronized.

Chris Date, one of the pioneers of distributed databases, has developed a number of rules that govern the operation of distributed databases. These include:

- Distributed databases must offer continuous operation
- The operation of a distributed database should appear identical to a non-distributed database to users and programmers
- The user should not be aware that portions of the database are replicated
- Portions of the database should be able to run on different types of servers with different operating systems.
- Queries should be able to be optimized by the DBMS in the same way that they are for non-distributed DBMS

One final consideration that arises from distributed databases is that of security. When many computers are connected to create a single database, each computer presents a security risk, and each must be equally well protected. If a single computer is compromised, the entire database may be at risk.

3.5.3 Financial Transaction Software

A number of different methods for performing financial transactions on the Internet have been tried. Some have been successful, and others have disappeared. Credit cards have become the de facto standard for handling financial transactions on the web, partly because of intense promotion on the part of the credit card companies, and partly because credit cards represented a familiar and trusted way of making payments that was perceived to be secure because of guarantees made by the credit card companies.

An example of a credit card-based system for the Internet is the CyberCash CashRegister System, which makes use of the infrastructure provided by the credit card companies to facilitate the provision of electronic payment services by Internet vendors. CashRegister provides Internet retailers with HTML code that processes credit card payments and securely forwards the information to CyberCash for verification. Encryption is accomplished using Secure Sockets Layer (SSL) or Transport Level Security (TLS). Once received by CyberCash, the credit card number of validated, a check is run to ensure that the credit card has not been stolen, and the card's credit limit is checked to ensure that it has not been exceeded. If the transaction is approved, the credit card is debited, and the money is transferred to the vendor's account. Finally, a confirmation is sent to the vendor, indicating that the transaction has been successfully processed, which allows the vendor to arrange for shipment of the goods purchased.

Although this system works well for one-time transactions, the effort involved in entering a credit card number, security codes, name and address information creates a barrier between client and vendor. Understandably, this is not popular with retail Internet vendors, since any barrier acts to reduce sales. To streamline the process of making online purchases, the credit card companies have developed the e-Wallet system. An example of this system can be found with Amazon.com's one-click ordering system (Figure 19). The system works by asking frequent clients of on-line stores to register with the store and provide their address and credit card information. This information is kept securely on file by the vendor, and the client's account is secured with a user name and a password. Once logged in, the client can simply click on an item to purchase it, and when the client logs out, the system confirms the items to be purchased. Since the client's address and credit card information are on file, the item can be shipped immediately, with no further involvement from the client.

The e-Wallet system is administered by the credit card companies (MasterCard and Visa), and is licensed to local banks. Obviously, as with any set of financial transactions that must be transmitted electronically, extensive use is made of security protocols such as SSL and TLS (Ince, 2004).

A similar system called PayPal has been developed to handle electronic payments. The only difference between PayPal and the e-Wallet system from the consumer's point of view is that PayPal is a two-way system. This means that payments can flow in either direction, so it is a more complete model for electronic financial transactions. Like the e-Wallet system, PayPal makes use of the infrastructure set up by the credit card companies, however, PayPal charges vendors directly for use of their system (Carr, 2007).

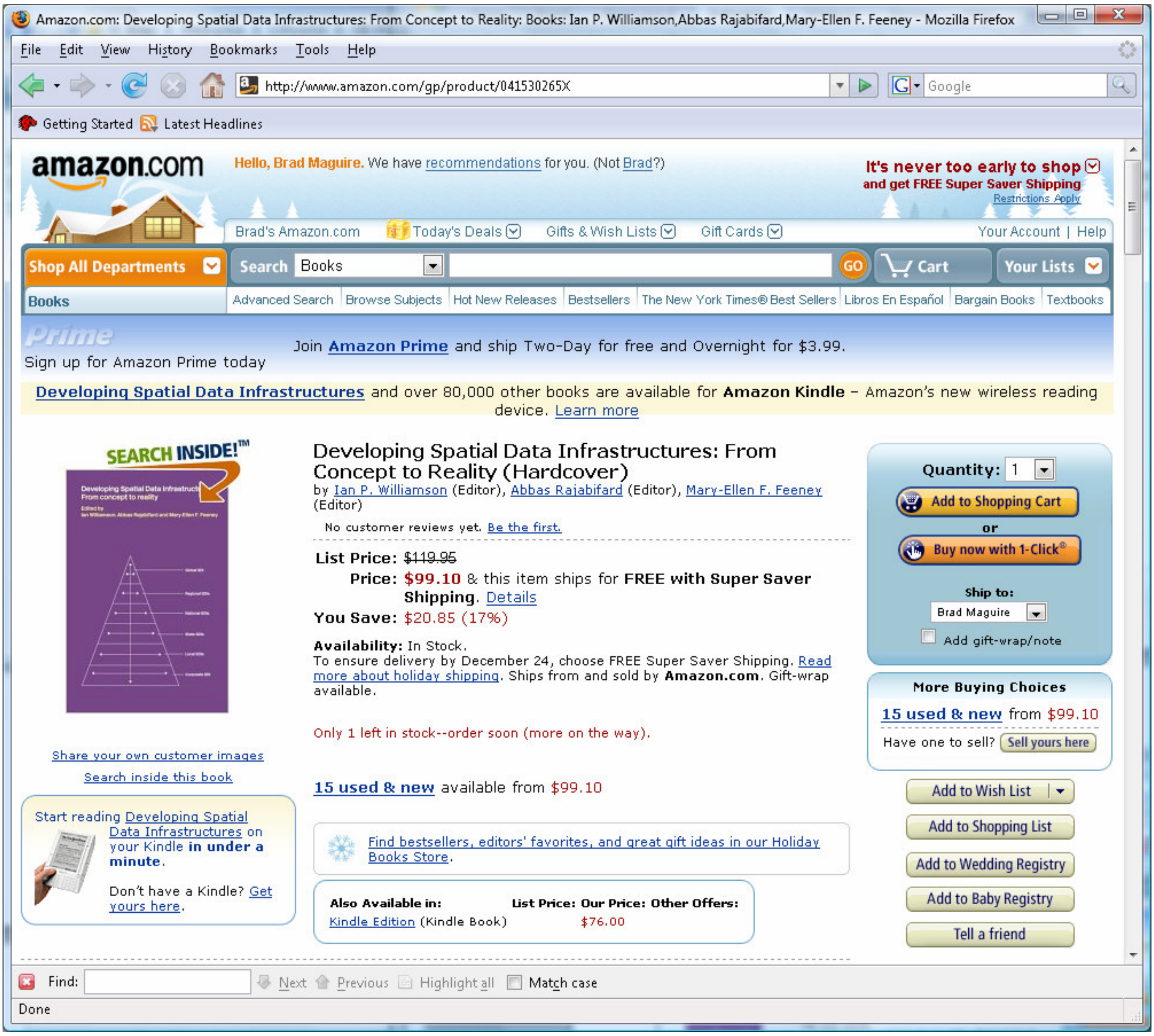

Figure 19. A web page from Amazon.com's portal, showing the e-Wallet and One-click ordering systems (Amber and Orange buttons on the right side of the page).

Another system called Clareon is designed for business-to-business financial transactions. Clareon is designed for larger, more formal transactions, and is designed to model the steps used in traditional business-to-business financial transactions. Using Clareon, the client and vendor first agree to make a transaction. The vendor sends the client an invoice electronically, after which the client sends authorization to Clareon to make the payment. This system makes use of digital signatures to ensure that the information was sent by an authorised representative of the client company. The system then makes appropriate debits and credits to perform the financial transaction, and sends conformation information to both the buyer and seller in a format that is compatible with each company's accounting systems. Thus, Clareon integrates well into existing business processes, and so is more likely to be adopted by many companies, since the system approximates how business was done on paper in the past (Ince, 2004).

3.5.4 Supporting GIS Software

When a request for data has been received, the data needs to be converted into the format requested by the data user. In the past, this has been a challenge, given the number of formats in which a particular data set can be presented. Fortunately, data conversion software is now available that can convert data between formats easily and automatically, once the conversion parameters have been defined between the different formats. This enables a GII to store data in a single format internally, and convert it “on the fly” to other formats. The Feature Manipulation Engine (FME) Spatial Direct server software by Safe Software is an example of this sort of universal GIS translation software. Spatial Direct permits translation to and from more than 200 raster and vector formats, and conversion into many different map projections. Other examples include Oracle Corporation’s Oracle 11g Spatial Database, and the CubeWerx CubeSERV package.

When a request for a formatted map is received, the GII must extract data and format it cartographically. To accomplish this, products such as the Environmental Systems Research Institute’s Arc Internet Map Server (ArcIMS) and AutoDesk’s MapGuide have been developed. These software packages can produce standard cartographic products from requested map layers and display these as graphics for clients who are using web browsers.

Image Processing will be accomplished using an image server program. An image server performs “on-the-fly” processing of images including radiometric and geometric correction and production of custom output products such as the Normalised Difference Vegetation Index (NDVI). This allows raw satellite imagery to be stored in the GII, and it can be processed to meet the needs of users when required.

Metadata Harvesting

In addition to supplying data to the GII, each contributing organization must also make metadata available so that the users of the GII are able to locate the data provided. Since these data are made available through the organization’s servers, the GII software must have the ability to access updated data and metadata, and make it available to all users through the main GII portal. The process of preparing the metadata for public accessibility is known as metadata harvesting.

Metadata harvesting consists of the discovery, validation, and collection of new metadata from contributing servers. The GII metadata harvesting software must identify the new metadata (and ignore metadata that has been previously harvested), examine its contents to ensure that required metadata fields have been populated, and then copy the validated metadata records to the GIS server.

A number of different methods have been created to allow metadata to be transferred to the GIS server, including Z39.50, Arc Internet Map Server (ArcIMS) metadata service, Web Accessible Folders (WAF) containing metadata in XML format, the Open Archive Initiative (OAI) metadata service, and the OpenGeospatial Consortium’s Catalogue Services for the Web (CSW) 2.0 format. These services involve the collection of the entire metadata record or only those fields that are required for data discovery from the GII Portal.

To supplement the metadata collected from the individual organizational servers, the GIS Server will append standard template data, such as legal statements, information on standard data elements (i.e. base map layers that are present on all maps), and standard definitions of terms (Creating and Publishing Metadata, 2007). One key to ensuring that data is discoverable is to have tools for handling the many variations on place names and feature types. To this end, an electronic thesaurus should be able to translate user queries using non-standard terms into standard keywords (such as the ISO 19115 Topic Categories) used by the data discovery tools (Nebert, 2004).

3.5.5 Portals

Portals are websites that collect data from many different sources and present them as a single web page with a unified style. The idea behind a portal is that it presents a single "doorway" to many Internet-based resources created by a company or organizations. Access to a portal is made available though a single User ID and password combination, which makes accessing data much easier than it would be if all of the applications made available through the portal were accessible separately.

A portal is presented to the user as a single web page with a number of frames (Figure 20). Each frame is actually the interface to a separate program possibly running on a separate computer, so in the context of a GII, one frame might display metadata for a map (coming from a metadata server), another might display pricing information (coming from financial transaction software) and a third might display a representation of the location of the data (coming from a map server).

Portals are a convenient way to allow users to interact with complex systems such as the Lithuanian GII, because they are so highly integrated. The contents, application, and processes work together to present the appearance of a cohesive whole. Behind the scenes, dozens of computers may be working to assemble and present the information requested. In the context of the Lithuanian GII, the portal provides and controls access to the many different servers that make up the GII, including data, map, image, database, metadata, and financial transaction servers.

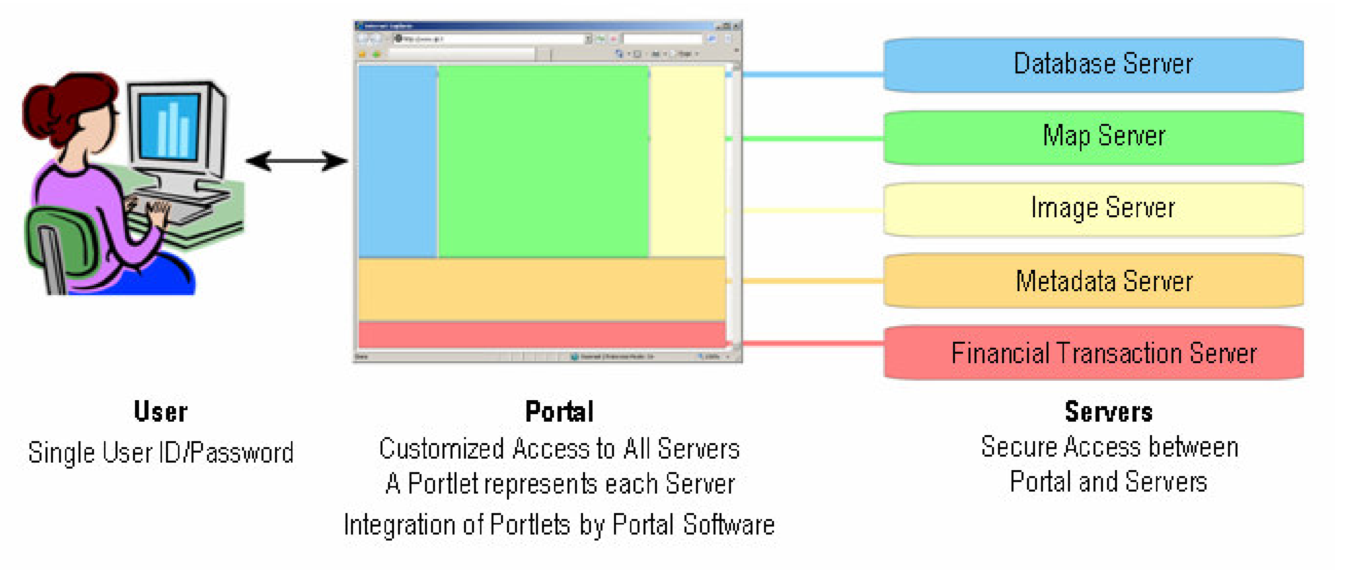

Figure 20. GII Portal Characteristics (Adapted from Hepper, 2005, p. 37)

Programmers are now creating the 4th generation of portals. The portal concept has been in existence since 1988, but new tools mean that the amount of custom programming required to create a portal is now much less than it was a number of years ago. Microsoft (SharePoint), IBM (WebSphere), SAP (NetWeaver), and BEA Systems (WebLogic) all produce commercial portal software; Apache produces open source reference software for portal design and experimentation (Root, 2005).

Each frame in a portal is controlled by a separate sub-program, called a portlet. Each portlet is an independent sub-program, which operates within a container provided by the portal application. By themselves, portlets cannot do anything, but when presented by the portal, they allow a great deal of information to be accessed efficiently. Portlets are also inherently a multi-platform solution. Not only can they run on desktop computers, but they can also run on cellular telephones and Personal Digital Assistants (PDAs), allowing wireless access to GII resources in the field. Portlets can be added or dropped from the portal, which allows users to customise the appearance and contents of the portal to meet their needs. For this to happen, portlets must be standardised.

There are a number of portal standards, the first being Java Specification Request (JSR) 168, Portlet Specification, which was released in 2002, and was followed by JSR 286, Portlet Specification 2.0, which was released in 2006. While JSR 168 provides basic portlet functionality, and is used widely, JSR 286 is just being adopted. This specification allows communication to occur between portlets, which allows the operation of the portlets to be tightly integrated, which enhances the illusion that the portal is a single program (Sun Microsystems, 2002, Sun Microsystems, 2006). Until recently, it has been quite difficult to integrate portlets into portals; this has required custom programming. The Web Services for Remote Portlets (WSRP) specification has been developed to allow portlets to be attached to portals easily, using only a few mouse clicks (OASIS, 2003).

Portal technology can go far beyond simply relaying information from servers to the user. Shi and Murthy discuss the concept of "Intelligent Portals", in which the portal software itself analyzes the user's use of the portal. Intelligent portals are distinguished from traditional portals in that the portal is separated from servers and applications by an information broker program, which actively chooses and filters the information presented to the client (Figure 21).

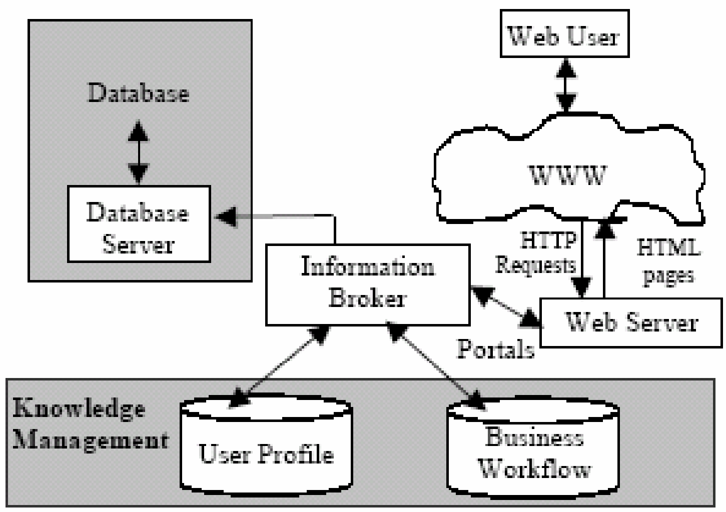

Figure 21. Intelligent Portal design, featuring an Information Broker program positioned between the portal software and the servers being accessed (Shi and Murthy, 2003, p. 47)

If, for example, the user favours a particular type of information, the portal can adapt itself so that that type of information is presented more easily to the user. This type of technology has found early adoption in commercial portals, such as those of retail bookseller Amazon.com (Figure). Amazon.com records the items examined in the on-line bookstore and uses the client's previous purchases to create a profile of the client's interests. This profile is matched against book keywords and books purchased by clients with similar profiles to produce a list of books that should be of interest to the client (Shi and Murthy, 2003). Of course, the Lithuanian GII need not be so commercial in its delivery of geospatial data to clients, but the capability of intelligent portals to help the user navigate his or her way to products of interest could be very helpful. Microsoft (SharePoint Portal Server), IBM (WebSphere and PowerBuilder), Oracle (Portal Development Kit), Viador (Viador), and Exsys (CORVID) all produce software to create intelligent portals. Unfortunately, unlike traditional portal offerings, this technology is new and there has been little standardization of interfaces and software to date (Shi and Murthy, 2003).

3.6 Conclusion

In this module, we looked at the hardware that makes GII possible. After quickly examining the hardware used by GIS, we looked at the computers and peripheral devices on which the GII software will run. We discussed how Server computers are designed for heavy, constant computing loads, and use redundant, industrial-strength components to reduce the chances of a computer failure significantly. We examined storage media in detail, focusing on optical and magnetic media, which are the two workhorses of today's computing systems. The advantages and disadvantages of tape and disc storage were examined, and the role of each was discussed in terms of keeping data secure. While discs are used for day-to-day data storage because of their high access speeds, tapes are used for backing up the discs because they can store greater quantities of data at a lower cost. Because backup techniques have to be able to recover files from several days to months previously, a great deal of data must be stored, and magnetic tape is currently the best type of media for this job.

The complexity of computer networks, as described in this module, is strongly related to the design and operation of GII. GII data and information, as we have seen in earlier lectures, may be freely accessible, restricted to paying customers, or restricted to certain government departments due to legal, privacy, open access, freedom of expression, educational and security concerns. Therefore, the design of the computer network for the Lithuanian GII both enables and constrains its ability to achieve its objectives. This design of the computer network must be related to political, economic, social and environmental contexts.

In this module, we have looked at the individual "Commercial off the shelf" software that is available for the construction of the Lithuanian GII. After logging in, the user of the GII sees a single portal which is made available to the user's web browser by a web server. The portal combines interfaces from a number of different servers, each of which runs different software. The custom software driving the portal sends requests to different servers that make up the GII, to produce a number of different output formats.

For users who are examining data, metadata and sample graphics can be produced by a GIS server to help the user determine whether the data files stored in the GII are appropriate for a particular purpose. Financial transaction software can handle electronic payments once the client has decided to purchase data. For purchased data, a data server can reformat and reproject GII data according to a customer's specifications, or a map server can generate graphical map products based on data requested by the user, delivering them in formats such as JPEG or TIFF files.

We have examined computer hardware, networks and software in this module. In the next module, we will discuss how all of the components that we described in this chapter are combined together to create a cohesive GII that behaves as a single object, and helps to make the Lithuanian GII a reality. Of course, a system of hardware, networks, and software are only so useful by themselves. In the second lecture of Module 4, we will examine the human issues of the GII as we discuss development, administration, and maintenance of the GII and the people and organizations that will control the GII.

Module Self-Study Questions

1. If you were building a Geographic Information Infrastructure, and needed to process remote sensing images, would you choose a Server or a Personal Computer? What characteristics of your chosen computer lead you to make this decision?
2. Using the Internet, do a search for "Server Computers" and have a look at the features offered by different computer vendors such as Sun Microsystems, Dell and Hewlett-Packard.
3. An effective System Administrator with a good backup plan becomes almost invisible to management. Only when a backup plan fails does the attention of management tend to become focused on System Administration. If you were a system administrator, what could you do to make management aware of the important job that you do?
4. What do you think would the key considerations in the design of computer network, which supports a GII dealing with sensitive, private data and information? In what ways would the GII and the computer network be synchronised to ensure effective management?
5. What do you think are the main differences between the network requirements of a desktop GIS user, and a large GII institution? What common problems will these different clients face?
6. Compare and contrast the different types of network topologies. How would the differences be relevant to GII institutions and users?
7. What problems do you think would arise if GII were based on a Virtual Private Network? How would you suggest the administrators should deal with the risk of viruses and Trojan Horse Programs?
8. A number of companies have attempted to create "micro-payment" systems for small Internet transactions (on the order of 10 centai) in the past. Virtually every attempt has failed. What is it about the Internet culture that rejects such schemes? How has an advertiser-based model been adopted in lieu of micro-payment schemes?
9. As we have seen in this module, the distributed computing model is extraordinarily flexible, allowing virtually any server-based software to be linked to create a portal. In what ways is this level of flexibility beneficial? How can it cause problems?

Suggested Readings

- "Server (computing)" Wikipedia article (http://en.wikipedia.org/wiki/Server_%28computing%29, Nov. 11, 2007).
- Chesher, Mary (2003). "The Storage Lifetime of Removable Media – Backup/Restore" Computer Technology Review, August, 2003 (http://findarticles.com/p/articles/mi_m0BRZ/is_8_23/ai_109665179/pg_1, Nov. 9, 2007)
- Shankland, Stephen (2007). "Windows Bumps Unix as top Server OS." *CNet News.com (http://www.news.com/2100-1016_3-6041804.html, Nov. 11, 2007)*
- Groot, R. and McLaughlin, J. 2000. Geospatial Data Infrastructure. Oxford University Press, Oxford.
- SANS Institute, 2006. TCP/IP and Tcpdump Pocket Reference Guide (http://www.sans.org/resources/tcpip.pdf, Dec. 6, 2007)

- Server GIS. (http://www.esri.com/software/arcgis/about/server_gis.html, Nov. 28, 2007)
- A Developers' Guide to the CGDI: Developing and publishing geographic information, data, and associated services (http://www.geoconnections.org/publications/Technical_Manual/CGDI_Technical_Manual_0204_e.pdf, Dec. 1, 2007)

Assignments

- Assignment 1: GII Network Components
- Assignment 4: Components of GII

References

- Chesher, Mary (2003). “The Storage Lifetime of Removable Media – Backup/Restore.” Computer Technology Review, August 2003 (http://findarticles.com/p/articles/mi_m0BRZ/is_8_23/ai_109665179/pg_1, Nov. 9, 2007)
- Cougias, Dorian J. (2003). The Backup Book: Disaster Recovery from Desktop to Data Center, 3rd Ed. Lecanto Fl: Schaser-Varton Books.
- “Fibre Channel” Wikipedia article (http://en.wikipedia.org/wiki/Fibre_Channel, Dec 9, 2007)
- Groot, Richard & McLaughlin, John (2000). Geospatial Data Infrastructure: Concepts, Cases, and Good Practise. Oxford: Oxford University Press.
- Khurshudov, Andrei (2001). The Essential Guide to Computer Data Storage. Upper Saddle River, NJ: Prentice Hall PTR.
- Masser, Ian (2005). GIS Worlds: Creating Spatial Data Infrastructures. Redlands, California, ESRI Press.
- Nebert, Douglas D. (ed.) (2004). Developing Spatial Data Infrastructures: The SDI Cookbook, Version 2.0 25 January 2004. Global Spatial Data Infrastructure
- “RAID” Wikipedia article (http://en.wikipedia.org/wiki/RAID, Nov. 12, 2007)
- Shankland, Stephen (2007). “Windows Bumps Unix as top Server OS.” CNet News.com (http://www.news.com/2100-1016_3-6041804.html, Nov. 11, 2007)
- “Server (computing)” Wikipedia article (http://en.wikipedia.org/wiki/Server_%28computing%29, Nov. 11, 2007).
- “Solid State Drive.” Wikipedia Article. (http://en.wikipedia.org/wiki/Solid_state_drive, Nov. 8, 2007)
- “Storage Area Network” Wikipedia Article (http://en.wikipedia.org/wiki/Storage_area_network, Dec. 9, 2007)
- Tate, Jon, Fabiano Lucchese, & Richard Moore (2006). Introduction to Storage Area Networks, 4th Ed. IBM International Technical Support Organisation (http://www.redbooks.ibm.com/Redbooks.nsf/ RedbookAbstracts/sg245470.html?Open, Dec 9, 2007)
- Treadway, Tom (2005). “Real-life RAID reliability” (http://storageadvisors.adaptec.com/2005/11/02/actual-reliability-calculations-for-raid/, Dec. 8, 2007)
- "InfiniBand" Wikipedia article (http://en.wikipedia.org/wiki/InfiniBand, 2007/12/09)
- Cisco Systems, 2006. Cisco Documentation (http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/ip.htm#wp4148, Dec 5, 2007)
- Groot, R. and McLaughlin, J. 2000. Geospatial Data Infrastructure. Oxford University Press, Oxford.
- SANS Institute, 2006. TCP/IP and Tcpdump Pocket Reference Guide (http://www.sans.org/resources/tcpip.pdf, Dec. 6, 2007)

- Carr, Houston H., & Charles A. Snyder (2007). Data Communications and Network Security. Boston: McGraw Hill Irwin.
- Creating and Publishing Metadata in Support of Geospatial One-Stop and the NSDI (http://www.geodata.gov/gos/metadata/CreatePublishMetadata.pdf, Nov. 28, 2007)
- Hepper, Stefan *et al*. (2005) *Portlets and Apache Portals.* Manning Publications (Available online at http://www.manning.com/hepper/, Nov. 7, 2007)
- Ince, David (2004). Developing Distributed and E-commerce Applications, 2nd Ed. Harlow, Essex: Pearson Education.
- Nebert, Douglas D., editor (2004). Developing Spatial Data Infrastructures: The SDI Cookbook, Version 2.0 25 January 2004. Global Spatial Data Infrastructure
- OASIS (2003) Web Services for Remote Portals Specification 1.0. (http://www.oasis-open.org/committees/download.php/3343/oasis-200304-wsrp-specification-1.0.pdf, Dec. 1, 2007)
- Root, Nate L. (2005) Say Goodbye to Portal Servers (https://www.microsoft.com/presspass/itanalyst/docs/Forrester_050324_Root_-_Goodbye_to_Portal_Server.pdf, Nov. 4, 2007)
- Shi, Nan Si & Murthy, V.K (2003). *Architectural Issues of Web-Enabled Electronic Business*. Hershey, PA: Idea Group Publishing.
- Sun Microsystems, 2002. JSR 168: Portlet Specification (http://www.jcp.org/en/jsr/detail?id=168, Nov. 3, 2007)
- Sun Microsystems, 2006. JSR 286: Portlet Specification 2.0 (http://www.jcp.org/en/jsr/detail?id=286 , Nov. 3, 2007)
- Wikipedia. List of TCP and UDP port numbers (http://en.wikipedia.org/wiki/List_of_TCP_and_UDP_port_numbers#Ports_0_to_1023, Nov. 24, 2007)
- Wikipedia. Web Server (http://en.wikipedia.org/wiki/Web_server, Nov. 24, 2007)

Terms Used

- Clareon
- Client-Side Scripting
- Commercial Off The Shelf (COTS) Software
- Commit
- Concurrency
- Database Management System (DBMS)
- Deadlock
- Distributed Database
- Downloading
- e-Wallet
- Extensible Markup Language (XML)
- Firewall
- Fragmentation
- Intelligent Portals
- Metadata Harvesting
- PayPal
- Portal
- Portlet
- Protocol
- Replication
- Rollback
- Secure Sockets Layer (SSL)
- Server-Side Scripting
- Stored Procedures
- Structured Query Language (SQL, “Sequel”)
- Transaction Management
- Transport Layer
- Transport Level Security (TLS)
- Web Server
- Web Services for Remote Portlets (WSRP)
- Active Networking
- Application Layer
- Application Layer Firewall
- Asynchronous Digital Subscriber Line (ADSL) Modem
- Authentication
- Authorization
- Bit
- Bridge
- Bus Network Topology
- Cable Modem
- Campus Area Network (CAN)
- Certificate Authority (CA)
- Checksum
- Client-Server Computing
- Complete Graph Network

- Data Security
- Default-Allow Firewall
- Default-Deny Firewall
- Demilitarised Zone (DMZ)
- Demodulation
- Denial of Service Attack (DoS)
- Digital Identity Certificate
- Distributed Denial of Service Attach (DDoS)
- Ethernet
- Fibre-Optics
- File Transfer Protocol (FTP)
- Firewall
- Full-Duplex
- Fully Connected Network Topology
- Heterogeneous Network
- Hierarchical Network
- Home Phoneline Networking Alliance (HomePNA)
- Hub
- Homogeneous Network
- Host-to-Network Layer
- Hybrid Network
- Hypertext Transfer Protocol (HTTP)
- Internet Layer
- Internet Service Provider (ISP)
- Layer
- Leased Line
- Line Network Topology
- Local Area Network (LAN)
- Mesh Network Topology
- Metropolitan Area Network (MAN)
- Modem
- Modulation
- Nessus
- Network Address Translation (NAT)
- Network File System (NFS)
- Network Layer Firewall
- Network Load Balancing
- Open Systems Interconnections (OSI) Model
- Packet
- Password
- Peer-to-Peer networking
- Perimeter Network
- Personal Area Network (PAN)
- Physical Security
- Point to Point Protocol (PPP)
- Proxy Server

- Public Key Infrastructure (PKI)
- Protocol
- Reverse Proxy Server
- Ring Network Topology
- Router
- Secure Sockets Layer (SSL)
- Security Administrator Tool for Analysing Networks (SATAN)
- Serial Line Internet Protocol (SLIP)
- Stack
- Star Network Topology
- Star-Bus Network Topology
- Star-Ring Network Topology
- Strong Password
- Switch
- Telnet
- Transmission Control Protocol (TCP)
- Transmission Control Protocol/Internet Protocol (TCP/IP)
- Transport Layer
- Transport Layer Security
- Tree Network Topology
- Universal Resource Locator (URL)
- User Datagram Protocol
- Virtual Private Network (VPN)
- Virtual Security
- Voice-Band Modem
- Wide Area Network (WAN)
- Wireless Local Area Network
- 9-track tape
- Antivirus Software
- Bit Error Rate (BER)
- BluRay
- Changed File Backups
- Compact Disc (CD)
- Compact Disc Read Only Memory (CD-ROM)
- Compact Flash Card
- Core Memory
- Data Mirroring
- Differential Backup
- Digital Versatile Disc (DVD)
- Digital Video Disc (DVD)
- Digital Video Disc Read Only Memory (DVD-ROM)
- Digitising Tablet (Digitiser)
- Ethernet
- Evolutionary Backups
- Fibre Channel
- File Archive Bit

- Flash Read Only Memory (Flash ROM)
- Floppy Disc
- Full Backup
- Helical Scan Tape Drive
- High Density DVD (HD_DVD)
- Hot Spare
- Hot Swappable
- HyperSCSI
- Incremental Backup
- InfiniBand (IB)
- iSCSI
- Local Area Network
- Mean Time Between Failure (MTBF)
- Mean Time to Data Loss (MTTDL)
- Memory Stick
- Network Attached Storage (NAS)
- Non-Volatile Solid State Media
- Normal Backups
- Parity Bit
- Parity RAID
- Plotter
- Portal
- Power Surge
- Recycle Backups
- Redundant Array of Independent Drives (RAID)
- Scanner
- Secure Digital Card
- Serial Advanced Technology Attachment (Serial ATA or SATA)
- Server Computers (Servers)
- Small Computer System Interface (SCSI)
- Storage Area Network (SAN)
- Storage Network
- Storage Server
- Striping
- Tape Autoloader
- Tape Library
- Tape Stacker
- Thumb Drive
- Undelete Command
- Uninterruptable Power Supply (UPS)

4 Establishing a GII

In this module, we will examine the overall structure of a GII. In Lecture 1, we will look at the technical components, and how all the parts of GII fit together to make the whole. In Lecture 2, we will examine the human element -- the organisations, procedures, and individual roles that make the technical component possible

Outline:
1. Design of GII
2. GII Development, Administration, and Maintenance

4.1 Design of GII

The previous module examined the hardware, network, and software components of GII as individual units; we will now examine how these units fit together to create the GII system. The design of the GII creates synergistic effects, so that the whole is greater than the sum of the parts. Putting all of the software, networking, and hardware components together in an intelligent fashion creates a system with some very powerful and unique capabilities. If done properly, the system created can be scaled up to allow it to serve more people, and additional capabilities can easily be added as the technology is developed. Thus, the GII behaves almost like a living organism that grows and adapts to its environment.

4.1.1 GII Architecture

Issues in Creating a GII

Creating a GII is more than assembling a group of Commercial off the Shelf (COTS) software packages and making them work on a group of networked servers. There are many ways that these packages can be interconnected, and the custom connecting software that makes this possible still forms a large portion of the total GII software. For this reason, despite the impression that the titles of books such as the *SDI Cookbook V 2.0* might give, there is no template that can be universally applied to GII development, only a broad set of recommendations that help the people implementing GIIs avoid making the mistakes of the past.

No GII has ever been created in a vacuum. Existing practices, standards, software, hardware, and data strongly influence the form of the GII that eventually results. Different agencies have different ways of viewing the world, and while these differences may not be apparent in day-to-day conversation, they may become apparent when attempting to combine spatial data produced by different agencies. Different agencies may employ different modelling schemes. This may be the result of conscious decisions based on a particular understanding of the way the world works, or it may be the result of software limitations, which force a particular method of interpretation onto the agency.

In addition to software, different operating systems and database management systems must be made to work together. Syntactic differences may result from different representations of objects, which may be a by-product of different data models being used. Objects may be also differ because of the map projection/coordinate system or the methods of data collection used.

Perhaps even more difficult than these technical issues are conceptual issues: different ministries may have different operational definitions for similar features. In English, for example, the words “wetland,” "swamp," and "marsh" all have very similar meanings, so different ministries may classify the same feature in different ways. A wetland is a generic term for any area that has a mixture of water and land; swamps are a more specific term for areas that are more water than land, and marshes represent areas that are more land than water. So if two maps showing wetlands overlap, one layer may show “swamps” and the other may show “marshes.” Even photo interpreters working for the same agency might disagree on which is which, particularly if there is about 50% land and 50% water. The GII must be able to map between different semantic systems such as these to produce consistent, meaningful results.

Another issue may involve schematic differences between the systems that are being merged in the GII. For example, because different agencies have different foci, one may treat a "swamp" and a "marsh" as separate features, whereas another agency might have a single "wetland" feature, with an attribute “type” which indicates the type of wetland that it is (Groot and McLaughlin, 2000). To translate between these two different schemas, the file may have to be altered to change the attribute “type” to a feature code, or vice-versa.

Interoperability

A GII is a distributed system composed of physical networking, system services, and application software. The physical networking connects many different varieties of computers together, and allows these computers to communicate with one another to form a single system. Server computers run many kinds of interconnected COTS software which provide services to the system. These services may include data storage, data access, applications, and presentation of data. Application servers provide specific processing tasks which are made available to users on the network, and presentation servers take the information provided by many application servers and combine them to present a cohesive result to users of the GII. The application servers may run custom application software, which ties together all of the COTS software.

A service running on a server should be a self-contained application. It should not rely on any other software, and should be able to describe itself to other services, which allows the services to be chained together to combine the capabilities. This self sufficiency means that services can be added, dropped, or replaced relatively easily without affecting the other services in the GII, and indeed without affecting the GII itself very much. The service model is quite different from the traditional GIS model, in which vendors compete to produce the most powerful, all-encompassing products. In a service model, each service is compact, powerful, and well behaved (Nebert, 2004).

4.1.2 GII Discussion Model

In this model, we will look at the design of the GII as a series of subsystems, each of which performs a group of related functions. The model proposed in this lecture is for discussion purposes only, and is not an endorsement of any particular technology, configuration, product, or software vendor. At the present time, there are many competing standards and products, and there are even more ways of connecting these together to create a GII.

The easiest way to look at the overall design of GII is to examine each subsystem that makes up the whole. Moving from the Internet to the inside of GII, we have a Security Subsystem composed of Firewalls, Proxy Servers, and Reverse Proxy Servers. Immediately inside the Security Subsystem is a User Management Subsystem which allows the GII to handle requests from the many users who wish to obtain data from the system. The core of the GII is the Data Warehousing and Formatting Subsystem. This subsystem acts as a Data Warehouse and stores the GII data in a single format, but is able to reformat the data into many different formats that are requested by the clients of the GII.

The Data Warehousing and Formatting Subsystem is connected through a Firewall and other security measures to the Internet on one side, and to the network of government computers on the other side. This network, which is already in existence, forms the Data Preparation Subsystem, and will be used to perform human-assisted data preparation. Because the Lithuanian GII will be a revenue-generating system, it also includes a Financial Subsystem which allows the GII to process financial transactions and provide accounting support for managers and accountants.

Security Subsystem

The purpose of the Security Subsystem is to prevent unauthorised access to valuable components of the GII that could be stolen or damaged. Of course, the most valuable component of the Lithuanian GII is the ability to access the data that are stored in the system, however, the software and hardware components as well as the data itself are also very valuable and should be protected from intentional or accidental damage.

The first line of defence in the Security Subsystem is the Proxy Server (Figure 1). Proxy Servers are computers that run special software to accept input from the Internet, filter and modify it, and route it to appropriate computers that are not directly connected to the Internet. In this way, it acts as a "broker" that permits appropriate communications from the Internet to a Local Area Network and vice versa.

Proxy servers serve two important security functions. First, they may be set up for user authentication, and second they take the raw input from the Internet and examine the requests to ensure that they are legitimate. In the case of GII, it is this second function that is most important. Although Proxy Servers may be set up to perform user authentication and authorization using passwords, such functions are unlikely to be used directly in the GII, since it is designed for public access. By accepting only particular kinds of server requests, the Proxy Server shields everything that is behind it from attack. Since many forms of Internet attack rely on malformed Uniform Resource Locator (URL) requests, having these filtered out provides a first line of defence against outside attack.

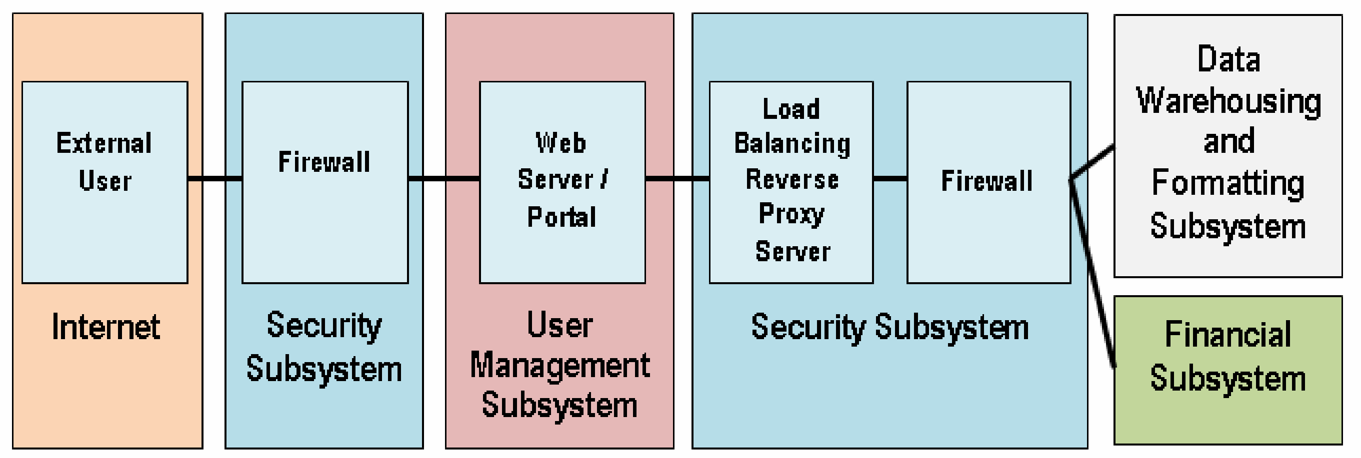

Figure 1. The "front-end" of the GII discussion model, including the Security, User Management, Data Warehousing and Formatting, and Financial Subsystems.

Reverse Proxy Servers (RPS) share many of the same characteristics as Proxy Servers, but offer increased security. The RPS acts as a single point of access, limiting the exposure of machines behind it. Only those servers that are on the access list in the RPS can be accessed; all other computers are completely blocked from outside access. The RPS is able to hide the identity of the machines behind it by remapping all URLs from an external format to an internal format, both for incoming and outgoing communications. In this way, it appears to users on the Internet that they are always dealing with the RPS, and that there are no other computers in the GII. Because malicious users are never able to directly identify the machines behind the RPS, it is nearly impossible for them to target a particular machine for attack.

By configuring the URLs on the inside of the Reverse Proxy Server in a particular manner, a default-deny Firewall can be set up to ignore all but the specially encoded URLs emanating from the RPS. Unlike most firewall configurations (default-allow), default-deny firewalls reject everything unless they match a particular pattern. This combination of an RPS and a Firewall creates a very strong barrier to unauthorised entry to the back-end computers in the GII.

The combination of Reverse Proxy Servers and Default-Deny Firewalls enables machines on a Local Area Network to be used both internally and from the Internet. Particular computers, such as those that provide geographical data to the GII, may be accessed by external users through the filter provided by the RPS, or by internal users, who are working in a trusted environment. Other machines on the internal network, such as those for the management of the GII, are only accessible by internal users. This configuration allows internal computers to be used effectively, and avoids the need for separate external computers to service the external users of the GII.

Reverse Proxy Servers can also perform load balancing. If more than one computer can fulfil a request, then the RPS can choose the machine which is least heavily loaded when it reformats URLs. The RPS can thus help to make the GII scalable, allowing additional servers to be added as demand for geographical data increases. Furthermore, a change to the list of available servers in the RPS allows them to be swapped out for maintenance without interrupting service (Nguyen, 2004).

User Management Subsystem

When an external user logs into the GII, the first thing that they will see is an introductory screen welcoming them to the Lithuanian GII, and allowing them to create a new account or login with an existing account. Account and password management will be handled by the Web server. Once the user has created a new account and has logged in, the GII Portal interface will be displayed, and the user can begin to browse for geographic information of his or her choice using the tools provided for data discovery.

Tools for Examining Metadata

A number of different mechanisms will be available to allow users to search for data, including searching by location, type of data, or date of data. Users may search by location using a number of different entry methods including searching by drawing an area on the map to define a geographic area, entering the bounding coordinates of the study area, or selecting place name key words. Keywords may also be used to choose the type of data required; in addition users may select a class of information from a hierarchical tree showing what is available in the GII. Selecting data by its date of creation or modification can be accomplished by indicating a beginning and ending date within which the requested data must fall.

The above model assumes that the search for data is entirely keyword driven, but the user may wish to identify an area of interest on a map, so the Geospatial Data Catalogue must provide a map interface to supplement the keyword search.

The process of data discovery is relatively simple and consists of four steps, but the details of the operations can be quite complex. These four steps for data discovery are:

1. The user fills out a form which identifies the location, type of data, or date of data
2. a search request is passed to the Catalogue Gateway
3. The query is submitted by the Catalogue Gateway to a number of registered catalogue servers
4. Each catalogue server replies with a list of all hyperlinks and instructions about how to access the data. URLs are provided on which the user can click to access the data repositories (Nebert, 2004).

Users may also wish to consult more detailed metadata files than are stored in the Geospatial Data Catalogue, so at this point they may wish to contact the provider of data to requests a supplementary metadata or two task questions about the quality of data or the utility of the data for a particular purpose.

In this system, we are assuming that individual layers of GIS data are available for purchase. Traditionally, map data has been sold “by the sheet,” and when you obtain data, you get all of the layers for a particular mapsheet. In order to make GIS data reasonably priced in the GII, however, it is better to sell data “by the layer and the sheet,” so that users only have to purchase the data that they require. This removes the disincentive of excessive cost, reduces the inclination to share data once it has been purchased, and means that a higher price can be charged for each layer than would otherwise be the case.

The Catalogue Services for the Web 2.0 (CSW 2.0) standard developed by the Open Geospatial Consortium is designed to standardize the discovery and exploration process for metadata. It defines the user interfaces that are used to discover, browse, and query metadata (Open Geospatial Consortium, 2007)

Being able to locate data requires that high quality metadata be available. Metadata acts like a catalogue in a library; if it has a standard format, it is reasonably easy to find information, but if the format is not standardized or if fields are missing, it can be impossible to find what you want. For this reason, there are a number of metadata standards available. The International Organisation for Standardisation (ISO) 19115 standard defines the layout of metadata records, with required and optional fields. ISO 19115 allows metadata levels to be nested, so that local metadata meshes with Global Metadata (Nebert, 2004). Other metadata standards that might be used are the ISO 15046-15 standard, the CEN (Comité Européen de Normalisation) Dublin Core metadata standard, and the Federal Geographic Data Committee (FGDC) metadata standard (Groot & McLaughlin, 2000).

Metadata plays three roles in the discovery of spatial data:

1. It documents the location of information
2. It documents the context and structure of information
3. It provides the user with detailed information on appropriate use of the data

Data Warehousing and Formatting Subsystem

The Data Warehousing and Formatting Subsystem, together with the Financial Subsystem, and the Data Preparation Subsystem form the "back end" of the Lithuanian GII (Figure 2). In this section, we discuss four components of the Data Warehousing and Formatting Subsystem, the DBMS Server, the GIS Data Server, the Map Server, the Image Server, and the Catalogue Server.

Three different kinds of data are stored in this subsystem and its components reflect this. Attribute data, in which data are represented as text or numbers in a table, are stored and distributed by the DBMS Server. Vector data, in which features are represented with points, lines, or polygons, are distributed by the GIS Data Server. Raster data, in which features are represented as pixels in a grid, are distributed by the Image Server and the Map Server. The Map Server integrates data from the DBMS Server, the GIS Data Server, and the Image Server to create new raster maps. The Catalogue Server harvests metadata from the DBMS Server, GIS Data Server, and Image Server, to create a catalogue of metadata that is stored on the server. Although these components have separate and distinct functions, commercial products blur the distinctions between them. In this section, however, we will discuss these components independently so that their functions can be explained better.

DBMS Server

The Database Management System (DBMS) Server is used to control the storage of GIS attribute tables and possibly graphical information. In the context of the Data Warehousing and Formatting

Subsystem, the DBMS serves as a general data storage and retrieval system which is under the control of the GIS data server. For users who do not require access to data in spatial form, the DBMS can be used to produce tabular reports on the data stored in the GII, without requiring the involvement of the GIS Data Server.

GIS Data Server

The GIS Data Server processes the raw spatial data stored in the GII, and makes it available in a GIS-ready format, or passes this data through to the Map Server for further processing to create a finished map for immediate use.

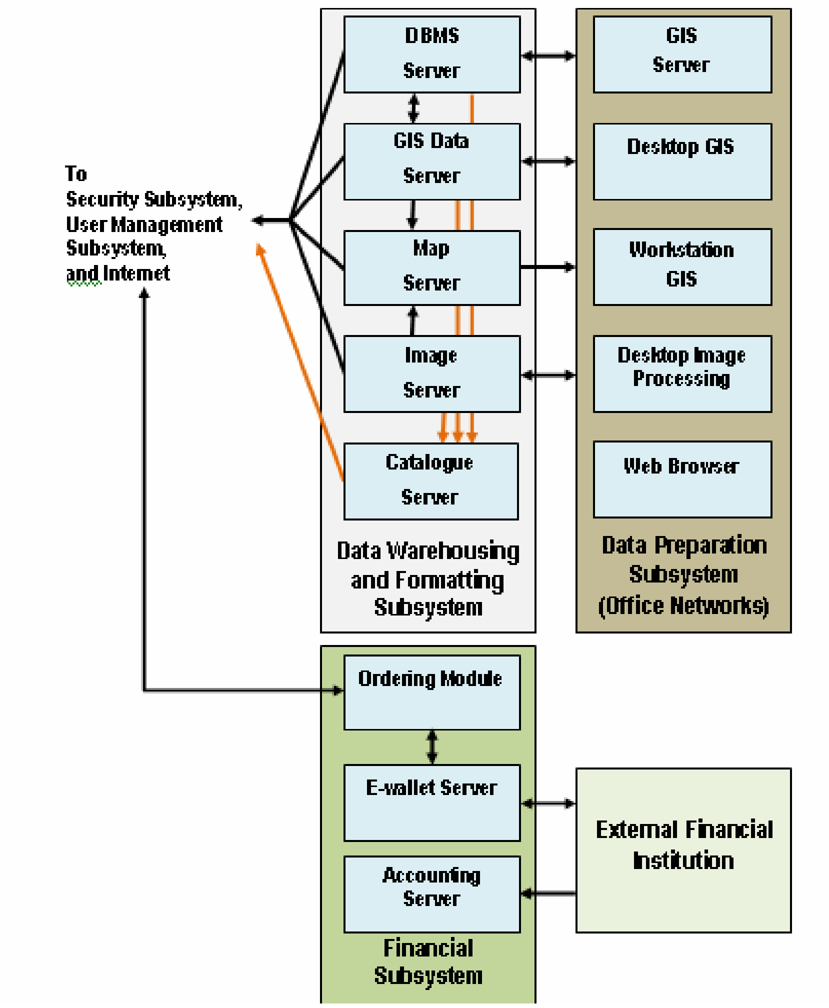

Figure 2. The "back-end" of the GII discussion model, consisting of the Data Warehousing and Formatting Subsystem, the Financial Subsystem, and the Data Preparation Subsystem. Black arrows represent the flow of data; red arrows represent the flow of metadata.

Within the GII, the GIS Data Server has relatively limited functionality, consisting primarily of changing the map projection or the format of geographic data to one of a number of common GIS data formats supported by the GII. The reason for this limited functionality is due in part to the desire to make the portal software in the User Management Subsystem "user friendly," which strongly encourages a limited set of functionality to be made available to the non-professional user, at least until they choose to use "advanced" functions. Professional users of the GIS Data Server, who are accessing it from the inside of the GII, will not be limited in the functionality that is available to them. Of course, with reduced limitations comes an increased requirement for GIS training to be able to use the software correctly and effectively.

Data Access and Delivery

Spatial data in a Geographic Information System is typically stored in one of two ways. Vector data use points, lines, and polygons to outline the boundaries of discrete features on the air surface. With raster data, features are represented with a regular grid of pixels.

In a vector dataset, a small feature such as a fire hydrant or a road sign would be represented as a single point, a road or river might be represented as a line feature, and a municipality or urban development zone might be represented as a polygon. All of these features are ultimately created from points. A feature can be represented as a single point, two or more points can be joined by a straight line segment to create a line, or three or more points can be joined by straight line segments to create a closed area, forming a polygon (Figure 3).

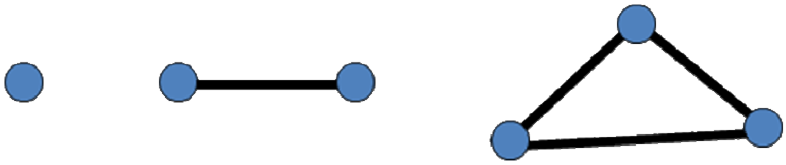

Figure 3. Point, minimal line, and minimal polygon features in a vector system.

Rasters may represent *continuous data*, which has no boundaries, *discrete data*, which has sharp divisions between values that represent boundaries, or images, such as those produced by a digital camera or a remote sensing satellite. Unlike vector data, where the boundaries can appear quite smooth if enough points are used, the boundaries of discrete features appear quite blocky (Figure 4).

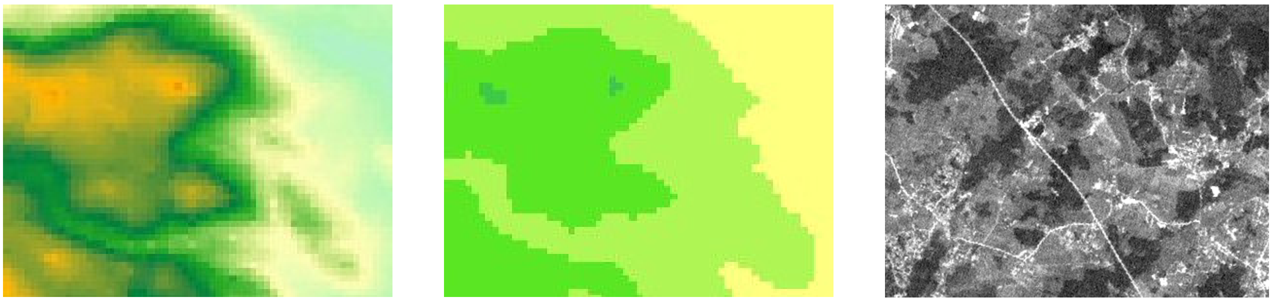

Figure 4. Continuous data (raw elevation values), discrete data (elevation classes), and image data (satellite image) represented in rasters.

The differences between vector and raster representation of data call for different tools within the GII to handle them. Vector data needs to be processed by the GIS Data Server before it can be delivered to a GIS or transferred to a Map Server for further processing. Raster data, in particular satellite images, may need to be pre-processed and classified by an Image Server before they can be delivered to the client or transferred to the Map Server.

Delivering GIS Files to the Client

There is a broad range of ways in which GIS data can be delivered to a client, ranging from the simple to the sophisticated. These simplest solutions require human intervention, for example printing out paper maps, or creating CDs or DVDs containing data for a client. More technically sophisticated solutions are able to deliver data to the client over the Internet entirely without human intervention.

For those requests that cannot be handled automatically, when an order is received from the Ordering Module and there is a record on the Accounting Server that payment has been made for the data, a human operator must process the request, package the product for delivery, then send it by courier to the client. Delivery times for this mode of operation may be a short as overnight, will more likely will require several days to process, format, and export the data, particularly if the request is quite complex.

For purchases of data that can be handled automatically, the GII is able to deliver the data immediately over the Internet, once it has been purchased. For simple requests, such as the delivery of GIS data in its native format, or even with a changed projection and data format, it should be possible for the GII to process the data immediately. In such a case, no human intervention is required, and money is simply deposited from the E-wallet server into the project bank account, and the Accounting Server is updated to record the transaction.

Electronic delivery may consist of placing the data in a password-protected location on the Internet for pickup by the client, the automatic delivery of the data to the client's computer, or even direct delivery into the client's GIS software. Electronic delivery for client pick up might be accomplished using a File Transfer Protocol (FTP) site, in which individual deliveries are password protected. A similar option would be to use the same Web Accessible Folder (WAF) mechanism that has been proposed for the transfer of metadata from government GIS servers to the GII. A slightly more sophisticated option would be to automatically packaged the data delivery as a single compressed file (.zip, .rar, .gzw etc.), and then provide the user with a button to download the file from the Web portal, or even to automatically trigger the file download to the client’s computer.

Proprietary and Open solutions also exist for the transfer of data directly into GIS software. These software solutions typically make use of the Extensible Markup Language (XML) which is a language that was developed from HTML (Hypertext Markup Language), but which allows virtually any feature in existence to be defined. XML is a standard developed by the World Wide Web Consortium (W3C). Every time a new group of features is defined using XML, a new data transfer language is defined. Thus, there are many data transfer languages now available that are based on XML. Some examples include:

- CML: Chemical Markup Language
- SBML: Systems Biology Markup Language
- RSS: Really Simple Syndication
- ArcXML: Arc Extensible Markup Language
- GML: Geography Markup Language

Because XML can be used to define virtually anything, it is important that users be able to understand features that are being encoded using the language. For this reason, XML includes its own metadata, which defines the features being described. In addition, XML is a text-based language, which means that it is human-readable. In theory, given a printout of XML code (which would be very long) and its metadata, it is possible to extract the information from the file without a computer, which means that information encoded in XML is not "locked up" with a single piece of software, and is more useful for archival purposes than binary files which may prove impossible to read if the original software is lost. The disadvantage to having a text-based language is that it tends to create very large files which take a great deal of time to transfer.

The Open Geospatial Consortium also produces a system which allows GIS data to be transferred directly to GIS software on the client's computer. The software uses an OGC standard called the Web Feature Service (WFS), which transfers data from source to destination using the Geography Markup Language (GML), which is a version of XML that is designed for the transfer of GIS data.

Although there has been much discussion of ways to compress GML to allow for the transfer of large data sets, this has not yet been implemented as part of GML (and may never be). One solution to this problem, suggested by Ron Lake, is to make use of CODECs (coder/decoders) such as ExpWay's BinXML to encode GML into binary format, which is much smaller before transmission, and then decode the binary format back into regular GML once it has reached the client (Lake *et al.*, 2004). This allows for rapid transmission speeds, but also retains the tight integration between WFS and the browser on which data are being displayed. Another, less sophisticated method is to simply transfer data in another more compressed format, but this relies on the ability of the user of the data to be able to download the file and make use of it in his or her software.

Because GML is designed to transfer data for use in a GIS, it is designed only for the transfer of spatial information, not for the transfer of the *representation* of that data. Information on how GML data should be represented may be transmitted by another XML-based language called the XML Transformation Language (XSLT). The separation of the geographic data from its representation allows a map "style" to be applied to the geographic data immediately. HTML has a similar capability, in which a stylesheet can be applied to immediately change the appearance and formatting of a webpage. In a similar fashion, different XSLT files can be applied to the GML data to alter the appearance of maps according to various needs. This capability makes it much easier for amateur users of GII data to obtain cartographic products that meet their needs.

The following is an example of a school on a map that is represented using GML:

```
<Feature   fid="142" featureType="school" >
   <Description>Balmoral Middle School</Description>>
   <Property Name="NumFloors" type="Integer" value="3"/>
   <Property Name="NumStudents" type="Integer" value="987"/>
       <Polygon  name="extent" srsName="epsg:27354">
            <LineString  name="extent" srsName="epsg:27354">
                <CData>
                  491888.999999459,5458045.99963358
                  491904.999999458,5458044.99963358
                  491908.999999462,5458064.99963358
                  491924.999999461,5458064.99963358
                  491925.999999462,5458079.99963359
                  491977.999999466,5458120.9996336
                  491953.999999466,5458017.99963357
                  </CData>
           </LineString>
      </Polygon>
</Feature>
```

Notice that this record includes attribute information (name of school, number of floors, number of students), information on the map projection (epsg:27354), and the vector outline of the polygon feature (enclosed between the <CData> and </CData> statements).

Galdos Systems (www.galdosinc.com) produces a product called Cartalinea which distributes GML data using WFS. Safe Software's FME Spatial Direct server is another alternative to reformat and prepare spatial data for delivery to GIS platforms. The Safe Software products excel at changing map projections and output formats. Whereas Cartalinea supports Open Geospatial standards, Spatial Direct outputs to a number of proprietary map viewers using plug-in software.

Image Server

The Image Server is a specialized server which processes remote sensing images to create products which are suitable for export to GIS or Image Processing systems, or which can be passed on to the Map Server as a layer in a custom map being produced by the GII. Even though only a small amount of remote sensing data may flow through the GII, and Image Server is required because of the amount of computer processing that is required for Remote Sensing images (see GII-04, Module 4 for further information). As with the GIS Data Server, the Image Server will be limited in functionality when users access it via the Internet, simply because of the limitations imposed by a "User Friendly" interface design. Professional users on the inside of the Firewall may choose to use the Image Server to its full capabilities when, for example, creating custom Remote Sensing products such as supervised classifications.

The ESRI Image Server software allows rapid transfer of remote sensing images to clients. Images may be stored in their raw format, or with any degree of processing. The system is able to process images immediately, and is able to perform image enhancement, orthorectification, pan sharpening (increasing the resolution of multispectral data by making use of higher-resolution panchromatic imagery) and mosaicking of raster images (ESRI, 2007b). Other options include products based on the Open Geospatial Consortium's Web Coverage[3] Specification (WCS), such as the CubeWerx CubeSERV Web Coverage Server (www.cubewerx.com).

Map Server

Unlike DBMS, GIS Data, or Image Servers, which provide unaltered or slightly modified data from for GII, Map Servers combine information from each of these types of servers to produce finalised maps in raster form. Output from the map server is most important to non-professional users of GII, who do not have GIS tools with which to produce their own custom maps.

A Map Server has three basic functions. First it must produce a map. In this context, we are talking about a cartographic rendering of one or more overlaid layers of spatial data. Although features in each layer need to be represented cartographically (i.e. points, lines, and polygons should be coloured and symbolised, and raster data should be coloured), there is no requirement that other cartographic elements such as scale bars, map legends, north arrows, neatlines etc. be represented. Second, a Map Server should be able to answer basic queries about the data set, such as "what is the name of a city," or "what is the area of the county." Third, a map server should be able to form a chain of geo-processing activities, which means that it should be able to indicate to other software systems the types of output products that it can produce (Nebert, 2004).

The process for generating a map with a Map Server is as follows:

1. The user uses data discovery tools to identify the features that are to be displayed on the map.
2. The locations of the required data and methods for accessing it are returned from the Global Metadata catalogue.
3. The user confirms the data to be used in the map, and the data are downloaded from the servers to the Map Server

[3] GIS users who are familiar with older versions of Arc/Info may find this terminology misleading – a “coverage” used to mean a layer of vector data, but the OGC uses the term “coverage” for raster data.

4. The Map Server processes the incoming data and renders it as a finished map according to the parameters provided by the user
5. The Map Server displays the rendered map to the user as a graphics file (i.e. JPEG) (Nebert, 2004)

The Arc Internet Map Server (ArcIMS) by the Environmental Systems Research Institute (ESRI) expands considerably upon the basic capabilities for displaying static maps described above. The ArcIMS package is designed to meet the needs of many different types of users, and consists of a series of modules that can be added or removed to provide many different capabilities. In addition to providing static maps, ArcIMS can distribute interactive maps, GIS data, and metadata. ArcIMS allows for the querying of interactive maps on-screen and the extraction of subsets of data from these maps. It also provides browser-based tools so that users can search metadata, and it also features a tool for data and metadata harvesting.

The ArcIMS Data Delivery Extension makes use of technology from Safe Software (producers of FME) to perform projection transformations and format conversions, as an extra cost option to ArcIMS (ESRI, 2007a). ArcIMS can display maps to both unmodified Web browsers and to browsers with added plug-in applications. Another commercial map server is the Autodesk MapGuide product, which offers similar capabilities to ArcIMS, through the use of a plug-in application that extends the capabilities of the Microsoft Internet Explorer web browser.

The Open Geospatial Consortium has also published an extensive set of specifications for Map Servers in their Web Map Server (WMS) 1.0 specification. The specification allows for the combination of standard format data sets in different projections on different servers, with the automatic conversion of projections to ensure that all the features line up. While the WMS specification provides for the visual overlay of data, it does not allow for GIS-type overlay operations such as unions or intersections.

The WMS 1.0 specification has been implemented in software in products produced by CubeWerx and other vendors. Like the ESRI product, CubeWerx combines features of the GIS Data Server, Image Server, and Map Server into a single product. The CubeWerx CubeSTOR data warehouse product transfers vector boundaries using the Web Feature Service (WFS, which uses GML), and raster features using the Web Coverage Specification (WCS). The data is symbolized and presented according to WMS specifications before being displayed by the CubeXPLOR portal (Figure 5).

The ESRI ArcIMS product allows the transfer of GIS data directly into ESRI products such as ArcGIS. ArcIMS transfers data to a client's computer using this ESRI's own XML-based language called ArcXML.

Unlike GML, which is a fairly "pure" language for transferring vector GIS information, ESRI's approach to XML is more practical in orientation, and is designed to transfer the additional information that is required by their software. In addition to geographic data, ArcXML includes information on map elements (graphical features such as scale bars, north arrows, neat lines, map legends and graticules), map services (what features are available from which sources), and website configuration. These additions support key capabilities of ArcIMS, such as the ability to distribute raw geographic data for use in GIS, completed maps, or metadata.

Catalogue Server

Arguably, the greatest challenge in the construction of a GII is not the collection of data, the construction of a data warehouse, or the ability to distribute the data to users over the Internet. It is the ability to locate appropriate data sets from a choice of thousands of possible data sets that reside in the GII. To be able to locate data sets, users must be able to browse and search through metadata.

Like many other components of the GII, the Geospatial Data Catalogue is composed of software that runs on a server and provides a service that can be utilized by other components of the GII operating on the network. The Catalogue Services provided by the Geospatial Data Catalogue are synonymous with many other terms used in the GII lexicon, including "Spatial Data Directory," "Clearinghouse," and the "Geospatial One-Stop Portal." All of these tools allow metadata to be discovered through keyword, time and location searches (Nebert, 2004).

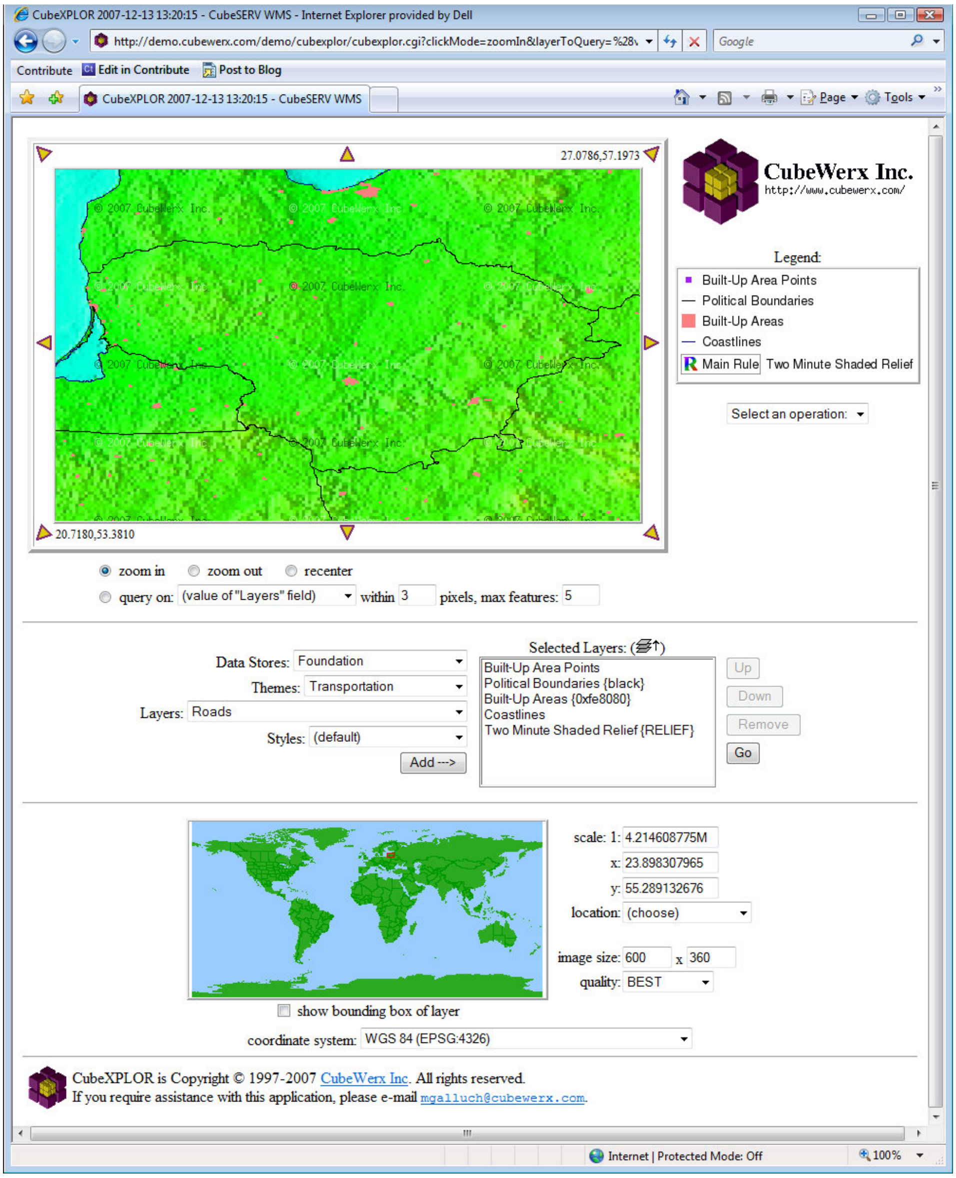

Figure 5. A browser window showing the CubeXPLOR portal, which uses WMS to display data from the CubeSTOR data warehouse. The vector boundaries for Lithuania were transferred using WFS/GML; the raster backdrop was delivered using WCS.

The term "Geospatial Data Catalogue" refers to a particular set of tools for identifying and revealing metadata, but it is not specific in terms of the way the tools are organised or the number of metadata records that may be accessed. A Geospatial Data Catalogue may refer to a single database on a single server, or it may refer to a very complex federated catalogue consisting of dozens of servers, each containing their own large catalogues of metadata. For larger catalogues, such as we might see for the Lithuanian GII, a number of different approaches may be used to combine the individual metadata holdings of the ministries that are contributing data and metadata to the GII.

Requests for metadata will be passed through the Reverse Proxy Server/Firewall combination before being passed on to the server that handles metadata requests. This ensures that only valid requests are sent on to the Data Warehousing and Formatting and Financial Subsystems, and that the computers that make up these systems remain protected from outside intrusion (Figure 6).

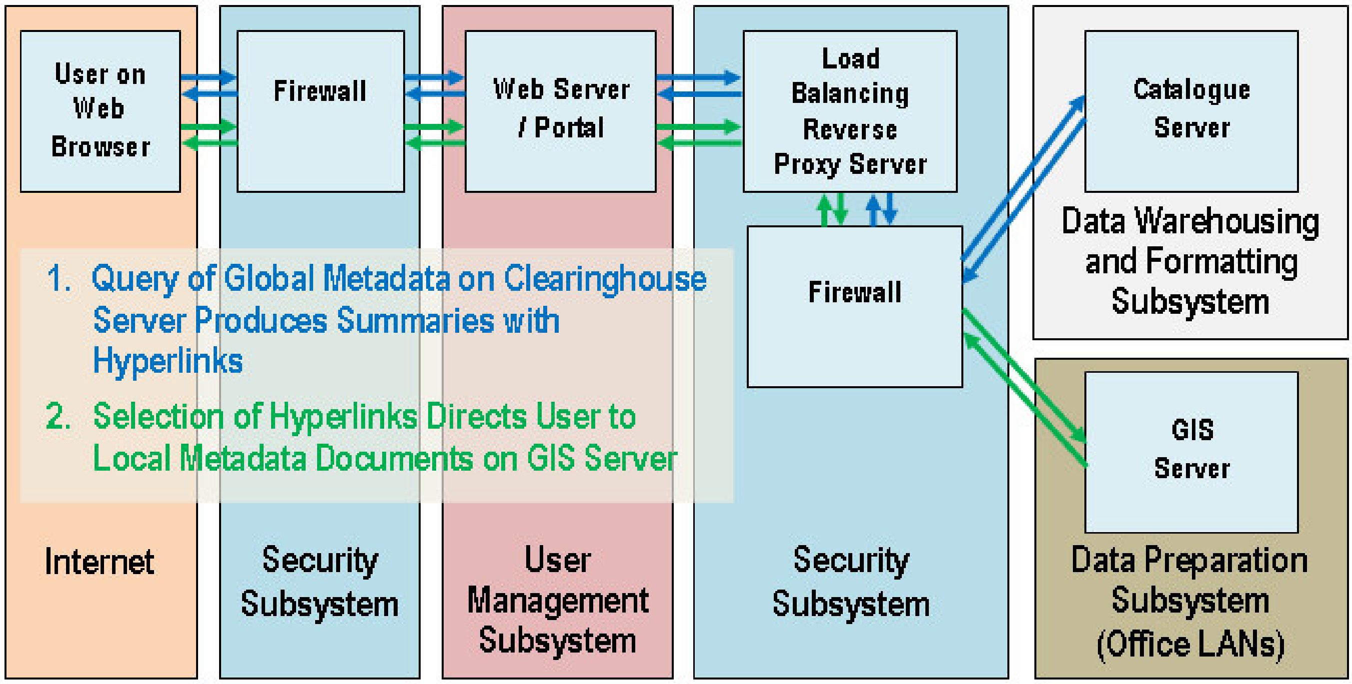

Figure 6. Catalogue Services returns Global Metadata, and then allows the User to access Local Metadata via hyperlinks (Based on Groot, & McLaughlin, 2000, fig. 9.5 p. 142).

In this discussion, we are assuming that the Geospatial Data Catalogue is an online, Internet-accessible utility. For users who do not have Internet access, another approach would be to publish the contents of the Geospatial Data Catalogue as a paper catalogue, or make it available on inexpensive media such as CD-ROMs (Nebert, 2004).

Metadata Harvesting

The harvesting of metadata may be as simple as copying the metadata from individual servers to a central location, may involve selection of particular records that are more valuable, or may even require the generalization of metadata into a more general description (probably involving manual editing of the records).

This gives rise to the concept of having two levels of metadata, Local Metadata, which is the raw metadata stored on ministry servers, and Global Metadata, which consists of the extracted or generalized metadata records that are centrally stored. The Local Metadata provides detailed information about the data sets stored on the ministry servers, whereas the global metadata contains a small subset of the information in the Local Metadata, and is used for identifying data sets for further consideration. In general, the higher the level of metadata, the less information is required (Groot & McLaughlin, 2000).

The procedure for harvesting metadata contains four steps:

1. Retrieval of all records from a metadata publishing location which is registered in the catalogue of metadata sites.
2. Validation of the records to ensure that all required metadata fields have been filled out[4]
3. Publishing of the metadata on the portal
4. Classification of the metadata, so that it can be reached using hyperlinks from lists of data available (Creating and Publishing Metadata).

The metadata might be submitted, either manually or automatically, to a central server for national distribution; such a system creates issues around the updating of data. The time lag that results from the update process may or may not be acceptable, depending on the needs of the GII users. This approach will require a system for the automated identification of the updated metadata and the harvesting of that data to a centralized metadata server for the GII (Nebert, 2004).

The transfer of metadata in either case may be facilitated by a number of different standards:

- The ISO 23950 Search and Retrieve Protocol (ANSI Z39.50) (OpenGeospatial Consortium Catalogue Surfaces Specification V.1.0)
- ArcIMS Metadata Service (Environmental Systems Research Institute [ESRI])
- Web Accessible Folders (WAF) (Microsoft)

Of course, the Web Accessible Folder does not specify the actual format of the metadata to be used, so metadata may be made available in commonly used formats such as XML or HTML.

The two-level catalogue is not the only model that has been proposed. Nebert (2004) divides metadata into three levels: Discovery Metadata, Exploration Metadata, and Exploitation Metadata. Discovery Metadata is a general description of data resources or data series, which serves to inform users of the existence of these resources and to provide a general description of what is contained in them. Exploration Metadata provides additional information that helps users determine whether the data will be suitable for a particular purpose, and provides contact information in case the potential user has further questions that are not answered by the metadata. Exploitation Metadata includes instructions on how to access, transfer, load, interpret, and apply data. It may contain a data dictionary, which describes all of the mapped elements as well as projection information. Another approach, made possible by the use of Reverse Proxy Servers and Firewalls, is to use a single repository of metadata in each ministry, which is accessible to both other governmental users and the general public through a highly mediated mechanism originating from the GII Portal. This approach carries with it security issues, although a proper security policy should make it virtually impossible for a user on the Internet to identify the servers involved or to be able to launch any sort of attack on them.

Data Preparation Subsystem

The Data Preparation Subsystem is a group of software packages that are used to prepare data for use in the GII. These software packages may include desktop GIS, workstation GIS, GIS servers, Web browsers, and image processing software. Much of the software that is used for the Data Preparation Subsystem may already be operational in the offices of government ministries. In its

[4] Nebert (2004) notes that it is easy to determine whether a column has been filled in, but difficult to assess the quality of the information that has been entered. Just because a metadata records passes validation does not mean that it is correct or useful.

role supporting the GII, the utilization of this software may change slightly; ministries may focus increasingly on the creation of reusable foundation and framework data, and the production of metadata to support reuse within the GII.

The Discussion Model assumes a high degree of integration between the GII and the Data Preparation Subsystem. The only barriers between the way that Internet users access computers in the Data Preparation Subsystem, and the way that GIS professionals in the government access the same machine are based on software restrictions. This enables the GII to be updated in a less restrictive fashion, and allows computer resources to be used more efficiently, but is inherently more risky than hardware-based security. If a hardware-based security system were imposed, there would be a physical separation between the GII and the computers in the Data Preparation Subsystem. This increased security comes at the cost of having to purchase additional server computers and having more limited access to the GII for updating data.

Financial Subsystem

The Financial Subsystem is a critical part of the GII, since it handles all of the electronic payments going into the system, and also contains the accounting software which is used to manage the system. The Ordering Module allows users to examine the prices for data in the GII and to place orders for that data. Once a price has been calculated, the E-wallet server is used to complete the financial transaction, and when the transaction has been completed, the server notifies the portals that it can release the data to the client, and also indicates to the Accounting Server that the transaction has taken place, so that the GII account can be credited for the amount of the purchase.

The Ordering Module would allow users to choose the data sets and the number of map sheets or the volume of data to be downloaded, and then would calculate a price for the client. The client would have the option of accepting the price quoted, or he or she could return to the selection of data sets and reduce or enlarge the amount of data to be purchased. At all stages, the Ordering Module would keep track of what is being purchased, and would display this for the client. Once the client has decided on the final data set for purchase, the cost of the data would be forwarded to the E-wallet system for processing.

The Discussion Model assumes that there will be large numbers of small purchases. The majority or the purchases would be from the general public over the Internet. For inter-governmental purchases, government credit cards can be used to handle ministry-to-ministry transactions. If, however, the ministry-to-ministry transactions are frequent and for large amounts, it might make more sense to install a business-to-business e-commerce system such as Clareon, which is designed for transactions of this type. However, if this is the case, it might be reasonable to reconsider data sharing agreements between government ministries, since the cost of maintaining a separate e-commerce application just for this purpose might be unwarranted.

4.1.3 Scaling Issues

One key advantage of the server-based distributed computing model proposed in the Discussion Model is the ease with which the GII can be scaled, either statically or dynamically. The server model allows multiple applications to run on a single server, so in the early stage of GII development, relatively little computer hardware is required. As demand grows, services can be moved on to new computers relatively easily, until each server runs a single application. As demand grows further, multiple servers can be set up to run the same application. A Load Balancing Proxy Server can be used to direct requests of a particular type to the machine with the least load. In this way, the GII, which initially ran on a handful of servers, can grow with increases in demand until hundreds of servers are involved in meeting the demand for services.

In addition to static changes in the configuration of the network, Load Balancing Proxy Servers can also be used to direct requests to fewer servers at particular times of the day. For example, during periods of peak demand, perhaps in the middle of the afternoon, all servers may be in use. In the middle of the night, when only a few requests are being processed, perhaps only one server will be in operation, with the rest of the servers being shut down or operating in a low-power mode. By ensuring that the numbers of servers in operation are proportional to the number of requests, response time and quality of service remain the same, no matter what time of day requests to the GII are made.

4.2 GII Development, Administration, and Maintenance

Module 4 of this course examines the computer and human resources required to establish a GII. In Lecture 1, we looked at issues of computer hardware, and how the components of the GII fit together to create a complete system. This lecture, will examine the human components of the GII, and how people can work together to make a GII possible.

4.2.1 GII Development

This part of Module 4 considers the development, administration and maintenance of GII. This topic is considered last, because it is fundamentally related to all the previous modules. This module is very important, because it considers issues which are better understood after examining the issues in the earlier modules, and the development, administration and maintenance of an effective GII is the ultimate objective of all GII developers. This involves a comprehensive knowledge of GII software and hardware, as was considered in Module 3.

Human issues are the main factors that enable, constrain and characterize the comparative development of GII in different countries. In Section 4.2.1, the development issues discussed include the strategies of the “first generation” of GII developers, which tended to focus on the development of national infrastructure of laws, policies, contacts and institutions, and those of the “second generation” of developers, who learned from the earlier developers, and tended to be more project and human focused. We will consider national, local and institutional policies and strategies in this section. We will conclude with an examination of the implementation procedures, software testing, the creation of metadata and the factors for financing of GII.

In Section 4.2.2, we examine administrative issues that affect the establishment of a GII. We consider the different roles of agencies that occupy leading and associated positions in the GII development hierarchy. We also examine funding issues, revenue sources and human resources issues such as training, recruitment and performance assessment.

Human factors, the mode of development, and the environment of GII influence the maintenance of the GII. In Section 4.2.3, we will examine the ongoing maintenance required for GII, including the updating of computer equipment and data.

By the end of this module, we will have tied together all the topics we considered in the all the lectures.

Development and Implementation Strategies

This section looks at the development and implementation strategies of GII, taking a comparative approach. The first topic looks at national level comparisons, between the first and second generation of GII developers, with examples taken from Europe, North America, Asia and Africa. This is followed by a discussion of the role of agencies in GII development, which is linked to the generational approaches. We then look at the implementation of GII strategies, and the factors to be considered in such implementation strategies. This is followed by a discussion of the methods of testing GII technology. These topics give a background to the next topic on GII maintenance strategies, by discussing the development issues that contribute to an effective maintenance strategy

First vs. Second Generation Approaches

The development and implementation strategies of GII are modelled on the approaches of different countries. This section examines the approaches of what may be termed the “first generation” of

GII developers, and compares these with those of the later “second generation” of developers. We examine certain key aspects of both modes of development. These include the existence of a “lead agency” that dominates the GII legally, politically, economically and/or technologically, and the associated, networked subsidiary agencies. We also examine the methods of GII testing inside and outside GII organizations.

The first generation of GII developers had several characteristics in common:

- The GII development had national coverage;
- The definition of GII included geographic information, spatial data, geospatial data and land information; and
- The GII included a coordinating mechanism for policy formulation and implementation.

The second generation of GII developers was characterised by:

- A following mode, rather than innovation, which may translate into the adoption of methods already evolved by innovators;
- The development of national level GIIs is usually part of an application to an existing problem, not an independent development; and
- A greater degree of extrapolation of method to other countries, due to the application to globally existing problems.

The second generation of GII users benefited from the experience, expertise and social capital of the first generation. These included:

- Direct learning from the first generation of developers;
- Development of partnerships in GII creation and management;
- A greater openness between the GII community members to the sharing of knowledge and data;
- A greater focus on organizational, research and conference materials;
- A larger number of forums for discussion and interaction, including workshops and conferences; and
- Publication of organizational, conference and research materials.

Table 1 illustrates some common differences and similarities between the first and second generations of GII developers.

Table 1. Similarities and Differences between First and Second Generation GII Developers (based on Williamson *et al.*, 2003)

Similarities and Differences	1st Generation	2nd Generation
Nature	Explicitly national	More flexible
Development Motivation	Integration of existing data	Links between people and data
Expected Outcomes	Linkage into a seamless database	Knowledge infrastructures, interoperable data and resources
Development Participants	Mainly data providers	Cross-sectoral (providers, integrators, users)
Funding/Resources	Mainly no specific or separate budget	Mostly include in national mapping program, or having separate budget
Driving Coordinating Agency	Mainly national mapping organizations	More independent organizational committees/partnership groups
Awareness	Low awareness at the beginning, more learning	More aware, knowing more about GII and its requirements
Capacity Building No of GII initiatives	Very low	Communities are more prepared to engage in ongoing activities
Number of GII Initiatives	Very limited	Many more
Development model	Mainly product based	More process based, or hybrid product-process
Relationship between levels (national/international)	Low	Much more
Measuring value of GII	Productivity, savings	Also holistic socio-cultural values

Examples of the First Generation of GII Developers

Australia and New Zealand. The Australia New Zealand Land Information Council was established as a national level organization that coordinates the collection, transfer and utilization of information on land use. This organization was situated in the Spatial Data Infrastructure Program of the Australian Surveying and Land Information Group. ANZLIC performed the functions of a lead agency "to lead the community in defining the components of the national spatial data infrastructure, the characteristics of those components, and provide a vehicle for the determination of national priorities and custodianship".

United States. An inter-agency Federal Geographic Data Committee (FGDC) was established in 1990 to coordinate "the development, use, sharing, and dissemination of surveying, mapping and related spatial data." Four years after the establishment of the FGDC, it was strengthened by Executive Order 12906, which was entitled "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure." The committee members included representatives of all government agencies interested in GIS. National level tasks included interagency coordination, the establishment of a national geospatial data clearinghouse and the development of a national digital data framework, including different levels of government and public and private agencies.

Portugal. Portugal was "the first European country that has an operational national geographic information infrastructure, fully distributed, based on the most recent data." In 1990, the

Portuguese government created the National System for Geographic Information (SNIG) and a National Centre for Geographic Information (CNIG) “to coordinate the integration of data at different levels of public administration and thus develop a National System of Geographic Information.” The CNIG developed the GIS market in Portugal, and coordinated two EC funded projects for the development of local municipal nodes for land use planning.

Holland. The National Council for Geographic Information (RAVI) was developed in 1992, out of the Dutch Council for Real Estate Information. It played an important role in the creation of digital core data sets, at the national level (1:10,000), and at larger scales for the municipal administration and public utility management. There is also a national geographic information clearinghouse project.

Korea. In 1995, the National Geographic Information System (NGIS) was set up for the development of digital spatial data bases and the standardization of geographic information. The national status of this organization was seen in the founding group, which included representatives from eleven ministries, and chaired by the vice-minister of Construction and Transportation.

Canada. The Canadian Council on Geomatics represents the provincial geomatics agencies. In 1996, this organization requested the Federal Interagency Committee to lead the guidance of federal and provincial governments, the public and private sectors, in the creation of the Canadian Geospatial Data Infrastructure (CGDI). Due to its national status, the head was the Assistant Deputy Minister of the Earth Science Sector of Natural Resources Canada. The Canadian Geospatial infrastructure identified five key themes for action:

1. The fostering of geospatial data access;
2. The provision of a foundation for framework data;
3. The harmonization of geospatial standards;
4. The encouragement of data sharing partnerships; and
5. The creation of a supportive policy environment facilitating the increased use of geospatial data.

The United Kingdom. The British National Geospatial Framework (BNGF) was founded in 1996, and was chaired by the Director General and Chief Executive of Ordnance survey of Great Britain. The board of the BNGF consisted of both public and private data users. The associated Association for Geographic Information also set up an advisory council, mostly data users, which worked with the government on geospatial data issues.

Examples of Second Generation GII Users

Colombia. The national data infrastructure in Colombia (Infraestructura Colombinana de Datos Espaciales, ICDE) has been described as “the set of policies, standards, organizations and technology working together to produce, share and use geographic information about Colombia in order to support national sustainable development.” The ICDE is a joint venture between several agencies, including the national mapping agency (IGAC), the national statistical agency (DANE), the national geology agency (Ingeominas) and the national oil company (Ecopetrol). Although developed without a clear, formal mandate from the government, the geospatial infrastructure is applied to issues concerning the natural resources of Colombia, especially biodiversity, mineral resources, natural hazards and socio-economic problems.

Hungary. The National Informatics Strategy was developed in 1996, and this contributed to the National Spatial Data Strategy (NSDS) in 1997. Other actors in this development were the Prime

Minister's office, other ministries, national institutions and the HUNGIS Foundation. The NSDS activities were initially located in the Prime Minister's office, but were later (2002) transferred to the Ministry of Information and Communications. For GIS functions, the most important agency within the NSDS is the Institute of Geodesy, Cartography and Remote Sensing (FOMI), operating within the Ministry of Agriculture and Regional Development. Most of the functions concern the application of GIS techniques to EU supported projects, relevant to Hungary's entry into the EU. Land registration is the major area of application.

South Africa. The National Spatial Information Framework was developed by the South African Department of Land Affairs in 1997. The objective of the project was to produce the parameters for a coherent GII characterised by the availability of relevant, reliable spatial information for planning, delivering services and the optimal allocation of resources. This work is supported by a Parliamentary bill promoting cooperation between public and private sectors involved in the development and implementation of a South Africa Spatial Data Infrastructure (SASDI). Within the Parliament, three subcommittees support the work of a Committee for Spatial Information, specializing in technical, liaison and policy issues.

Phased Implementation Strategies

A GII implementation plan must address the following technical, financial, and institutional considerations:

- System acquisition methods;
- Data requirements;
- Database design;
- Initial data loading requirements;
- System installation methods and timetable;
- System life cycle and replacement costs;
- Day-to-day operating procedures;
- Staffing requirements;
- User training; and
- Application development.

These elements may be developed in a phased implementation, in which feedback and strategic changes are closely monitored in each phase. Common problems that may justify a phased implementation include:

- The failure to identify and involve all users, i.e. operations, management, and policy levels of the organization.
- The failure to match GIS capability and needs. With a wide range of GIS hardware and software choices, it is challenging to make the right choice for the success of a GIS implementation.
- The failure to identify total costs, as although GIS acquisition cost may be easy to assess, it may be a small percentage of the total cost of GII implementation. Ongoing costs include hardware and software maintenance, staffing, system administration, initial data loading, data updating, custom programming, and consulting fees.
- The failure to conduct a pilot study, giving information on database design, data loading and maintenance, and day-to-day operations, with estimation of the operational requirements.

- The failure to consider technology transfer, including the training and support for on-going learning, for in-house staff, as well as for new personnel.

The following phases might be considered and implementing a new GII:

1. **Creating an awareness**, including the education of staff, provision of technical workshops (top-down approach), management workshops (bottom-up approach), and the identification of existing organizational problems which may justify the development of a GII. Issues may include:
   - Poorly maintained or out-of-date spatial information;
   - Spatial data not being recorded or stored in a standard way;
   - Spatial data not being defined in a consistent manner;
   - Spatial data not being shared between organizational departments;
   - Inadequate data retrieval and manipulation capabilities for existing needs;
   - New demands beyond the capabilities of existing information systems.
2. **Identifying system, organization and network requirements**, based on a user needs analysis that identifies users of a system and all information products required by those users. It may also include a prioritization of the information products and the data requirements of those products.
3. **The translation of organizational needs into functional requirements,** which may result in a set of processing functions, system capabilities, and hardware requirements.
4. **System evaluations**, where evaluating alternative hardware and software solutions may be conducted in several stages. Candidate systems are identified, and information to support this process is acquired through demonstrations and information sources. There may be short listing of candidate systems, on a low level assessment, followed by high level assessments based on the functional requirements identified in the previous phase. The assessment takes into account organization priorities and functional translation. There may be a standard benchmark to determine the system of choice.
5. **Justifying the system acquisition**, with a cost-benefit analysis as the GII should provide substantial benefits over expected costs. Intangible benefits are also considered, with an evaluation of other requirements such as data base development requirements, technological needs, maintenance, training, and organizational requirements and reclassification of existing job descriptions for employed users and networked collaborators.
6. **The system acquisition and start up**, including pilot projects, to detect wastage, and to generate some early results to justify the application, especially in the case of a complex project application.
7. **The operational phase,** involving the ongoing maintenance, application, and development of the GII, and including the responsibility for the system and liability. There must be security and transaction control mechanisms to support the system, with a systematic approach to the management of hardware, software, and data.

Implementation and Testing

Although the GII is constructed from Commercial Off The Shelf (COTS) components, which have theoretically been thoroughly tested before being commercially released, the integration of these components into a GII creates issues around how the software and hardware functions together.

Testing proceeds from internal users to external users and from smaller components to larger assemblies.

Testing by User Class

Testing by User Class proceeds from internal testers to specially selected external testers. Beginning with internal testers, who understand the software well, provides a buffer which enables large (and potentially embarrassing) software errors to be removed before they can be seen by external users. This leaves only small, relatively insignificant errors behind for external testers to find. Because the software has been relatively thoroughly tested before being released to external testers, there should be very few bugs present; these bugs are by definition hard to find, and require a large number of individuals to thoroughly test all possible permutations of software commands that might lead to these bugs. Only by using a large number of external testers is it possible to reveal these relatively insignificant errors.

The combination of internal (alpha) testers and external (beta) testers ensures that the final product is well tested and exhibits few errors. Alpha testing is the testing of a preliminary software product. The software may include only the core functions (acceptance of inputs and generation of outputs), and not include the full functions required for the application. Typically, alpha testing includes Unit, Integration, and System testing (see below). Beta testing is the last stage of testing, occurring after alpha testing when the product (versions of the software, termed beta versions) is made visible to specially selected outside users. In some cases, the beta versions are given open public access, for increased feedback.

Unit, Integration, and System Testing

Despite the fact that the software and hardware components should already have been tested, these components will have to be individually unit tested to ensure that they function as advertised. This is termed "unit testing."

Once individual hardware and software components have been successfully unit tested, they should be tested as a group in "integration testing." Integration testing examines how units function together in a larger group. For example, using the subsystems described in the previous lecture, integration testing might be performed on the Data Warehousing and Formatting Subsystem, the User Management Subsystem, the Security Subsystem, the Data Preparation Subsystem, or the Financial Subsystem.

Once all of the subsystems have been thoroughly tested, the entire GII is tested, in a phase called "system testing." System testing ensures that all of the subsystems work together correctly, ensuring that the entire GII functions as it is supposed to.

Going Live

After testing is complete, the system is made available for public use. This is termed "going live." Despite all of the testing, it is impossible to fully simulate real-world conditions. Since people can be quick to form opinions, it is critical that the system work well from a user’s perspective when it goes live. For the developers of the system, the first few hours after a system goes live can be stressful. Many factors affect the opinion that the users have of the GII; few of these are under the control of the development team. These factors include:

- The complexity of the organisation and network it must serve;
- The project applications;
- The amount of time spent by users with hands-on access;

- The skills, aptitude and motivation of the users;
- The commitment and priority of the organization and management;
- The availability of relevant data;
- The choice of software and hardware platforms, across networks, nationally and internationally; and
- The effectiveness of any feedback, and the flexibility of adaptive management.

4.2.2 GII Administration

GII, like GIS, cannot exist without people to operate and maintain it. Although a single individual might be able to operate and maintain a GIS, the scope of GII is much too large for a single individual to keep it running. The number of different skills involved, and the amount of work required mandates an institutional structure for the design and maintenance of the GII.

A number of organisations may be involved with the GII, and these range from groups that are highly involved with the design and maintenance, to those that are only marginally involved. Because so many individuals and groups are involved, it is impossible to create a GII without significant financial resources; thus, part of the institutional response is the acquisition and maintenance of funding for the GII. Another concern is ensuring that sufficient numbers of well-trained staff are available to maintain the GII. Organisations must address issues of staff training in order to keep the GII running.

Lead Agency

A Geographic Information Infrastructure requires an institution to be in charge of coordinating the actions regarding the development and operation of the networked structure. The Lead Agency is usually an institution having a mandate related to geospatial data management, which links with other organizations mentioned below, administratively facilitates the functioning of the GII, providing office space and related facilities for its operation, managing its resources, undertaking the networking management functions and linkages with the other GII initiatives. As we have seen above and in the earlier lectures, the Lead Agency in GII is often a government ministry, state organization or an empowered private company. The Lead Agency is the prime mover of GII, and is frequently networked with other, subsidiary agencies. In all the examples of first generation GII developers described above, the Lead Agencies were located in government departments. The power of such Lead Agencies derives from the legal and administrative powers granted by the supporting governments. The second generation GII developers were similar in this respect, but there was generally a greater focus on specific resource activities.

Subsidiary Agencies and Groups

In addition to the Lead Agency, there are a number of Subsidiary Agencies and Groups involved with a GII. These include:

- Central and local government organizations;
- Information traders and publishers;
- Hardware and software vendors;
- Conglomerates;
- Nongovernmental organizations; and

- Educational Institutions and Academics

The public central and local government organizations include survey and mapping agencies, land and property related organizations, and statistical research and environmental science organizations. Such government agencies are usually both producers and consumers of geospatial data and information and comprise part of the network of key agencies. Networks linked to such organizations include specialised government laboratories and other organizations collaborating in both geospatial data production and consumption. In countries where public utilities (for example electricity, gas and water) are organised by such government agencies, geospatial data networks are linked to the provision of such services. In other countries private companies perform these functions. Other private companies may be included in the GII network as subsidiary agencies. These include banks, insurance, real estate and retail companies.

Information traders and publishers may use remote sensing and GIS techniques to derive information on both social and natural issues. Such organizations, frequently privately run businesses, may also be government owned research or information agencies. They may network with government agencies as consultants or partners in GII applications. In such cases information exchange may be supported by government legislation and foreign contacts.

Hardware and software companies produce, buy, sell and manage the technology for GII development. As with other organizations, these agencies may be publicly or privately owned. As such technology is necessary for GII development, all GII networks depend on such organizations. They may vary greatly in size, from large companies such as ESRI to smaller retail stores.

Conglomerates are usually large companies with divisions that are seemingly unrelated to each other. This may be justified, as it may allow capital to be allocated more efficiently. A relevant, hypothetical example would be a conglomerate that consists of a hardware store and a software company, where the hardware store has high cash flow, but few profitable investment opportunities, and the software company has low cash flow, but good investment prospects. When the two companies are combined, the cash from the hardware store can be used to make profitable investments that would not be made by the software company. In this way, conglomerates are important components for GII networks, especially in the second generation approach where diverse organizations must combine to achieve project goals.

Non-governmental organizations may be business concerns or non-profit organizations. In either case, they offer assistance to larger GII institutions in the implementation of geospatial projects. This is especially the case for the project-focused second generation GII developers. NGOs are key facilitators of the social and environmental development strategies of many developing nations, working in partnership with the government and in some cases, business concerns. This requires networks sharing GII, where such applications may be useful.

Academics are usually based at research institutions, teaching institutions or universities. These provide inputs in the form of software (for example the IDRISI GIS software developed in Clark University in the United States) and also research information on the development and implementation strategies of GII. Academic input may be from Departments of Geography, Environmental Science/Planning, Engineering, Computer Science and/or Statistics. Considering the increased social and managerial issues in second generation GII developments, inputs may also come from Departments of Sociology, Economics, Management and even Gender Studies. Universities may work in partnership with GII networks and organizations such as those listed above, and may also conduct private, funded research useful for later application.

Funding

Funding for GII development may be derived from public sources (e.g. taxes or export revenue) and/or private sources (e.g. business profits). Public funding requires justification for the socio-economic value of GII development. As we have seen in this section, GII development has been difficult in some cases, as governments must be convinced of the practical long-term value of GII development, for large investments in money, time and personnel to occur. In the first generation of GII developers this was frequently an issue, but the justification for GII development was based on arguments for greater efficiency in storage, analysis and acquisition of geospatial information, faster application of this information to real world problems, and financial savings resulting from these advantages.

In the second generation of developers, social and project gains were emphasised to a greater extent, this being the justification for public expenditure. In particular, GII was applicable to the key socio-environmental and land based problems facing these poorer countries, the technology allowed more interaction between developing and developed countries, and international exchanges were beneficial to local and national development.

Cost recovery models are based on the financial benefits of GII development, which offset the costs from a cost benefit perspective, as described above. Cost recovery models are based on the three factors. First, GII is an expensive, highly technical and managerially intensive system, involving both high costs and strong potential benefits. Second, increased GII applicability, accessibility and availability lead to increased cost recovery. Third effective management of the GII increases the returns and benefits.

Benefits from GII may also be seen to provide the justification for tax support of their development. However, governments may still wish to obtain operational revenues. GII, in producing value added products, provides potential financial revenue to justify the initial investment of tax money. This model encourages the cooperation of private and public organizations, in mixed service and marketing ventures with GII. It also encourages governments to look at GII, as a mechanism for economic development, rather than as a costly liability. Finally, GII facilitates international cooperation, and leads to increased international status for those nations that develop one.

Staffing Issues

In order to have an effective GII, well-qualified individuals need to be involved in the design, administration, and maintenance of the system. An effective Human Resources strategy, which involves the recruitment, compensation, education, and assessment of staff members, is required.

In order to obtain and retain an effective step for the GII, there must be an effective recruitment and compensation program. Recruitment needs to focus on advertising job vacancies in appropriate venues to reach the right audience. For technical skills, advertising at job fairs and local universities might be warranted, whereas in order to obtain managerial skills, advertisements might be directed towards professional magazines. Fortunately, GIS and GII are of interest to many people, so it is not generally difficult to obtain a list of qualified individuals to interview for job openings. Retention of employees, once hired, is more dependent on effective compensation, both monetary and non-monetary to ensure that employees are happy, and that their families are well taken care of so they can concentrate on their work. Regular review of employee compensation is also important to ensure that key employees are not lured away to higher-paying jobs.

As we saw in Module 2, training of staff involves several key issues. An effective educational program needs to consider the following facets of training:

- Targets to be achieved;

- Content of training materials;
- The order of presentation;
- The speed at which material is presented;
- The opportunity for repetition of material;
- The form of practice to be provided; and
- Methods of testing students.

In order to effectively train GII users, it is important to understand the mental model of users and experts. The effectiveness of a GII may depend on the ability of people to learn, manage, teach and develop the required methods. Aspects of the mental model required of users therefore include:

- Problem definition. Relevant real-world phenomena must be classified, linked into a problem or objective, to allow the rational design of a GII, with execution and evaluation.
- Defining required system functions. The design of the system functionality is adapted or selected to resolve the problem.
- User interface design. The means by which the operator connects with the system, including the hardware and software.
- Visualisation design. This includes information held within the system, describing the real world phenomena and the results of GIS analysis and modelling procedures which are displayed by the user through visual representations.
- Usability testing. Methods used to evaluate the ease with which the system can be used.
- Instruction and testing. The formal and informal help required by the user for the efficient operation of the system.

Staff assessment should be carried out routinely to help employees identify and focus on educational opportunities that are appropriate and can help them improve job performance. Such assessments also improve formal communication between employer and employee, which helps to maintain the efficient functioning of the organisations which support GII.

4.2.3 GII Maintenance

The maintenance of GII data requires frequent updating, financial management, hardware development, and monitoring and training of specialised and generalist staff. As we have seen in this and earlier sections, the functions of GII vary with the size of the networks, functions of the components, and the political, legal, economic and physical environment of the GII institutions.

Ongoing Maintenance

This requires awareness of both human and technical issues. A formal training curriculum is required to ensure that time is dedicated to learning the technology properly, including well defined milestones and adequate time and training mechanisms, and a focused and properly trained operations staff. The threshold point of the learning curve is typically around the two year time frame, dependent on the ability of the organization to establish a well defined and structured implementation plan that affords appropriate training and resources for technical staff. After data models are well understood and sufficient data compilation has been completed the learning curve accelerates. Once a formal application development environment is established and user needs are well defined, an infrastructure exists for effective application of the technology. Building operational applications based on formal functional specifications will result in continued accelerated learning.

In the productivity curve, the establishment of a formal implementation strategy will help to ensure that realistic expectations are met. The availability of quality data supplemented by a planned implementation strategy is the cornerstone for achieving a productive and successful GII. Many organizations realise a loss in overall operational productivity over the short term while the GIS platforms are being installed, staff is trained, the learning curve is initiated, and data are being captured.

Depending on the unique circumstances of the implementation process, the status of data compilation, and the organizational climate, increased productivity is normally reached between the second and fifth year of implementation. This is identified by the threshold point. Again, this is dependent on a variety of factors including:

- The skills and experience of the staff involved;
- The priority and commitment of the organization;
- The implementation strategy; and
- The status of data compilation.

The primary issue with implementing GII is achieving the threshold point of increased productivity in the shortest possible time frame. In other words, we are trying to minimise the time during which a decrease in productivity occurs. Of course, the issue of productivity is typically of greatest concern with private industry, e.g. forestry companies, whose fundamental existence is based on earning a profit. Nonetheless, the significant investment in hardware/software, data, and training necessitates that a structured approach be utilised to achieve the threshold point in the shortest possible time frame. GII development based on well defined user needs and priorities is more likely to succeed than without these. A major pitfall of most GIS installations is the lack of well defined user needs on which to base the GIS acquisition and implementation.

Updating and Replacement of Computer Equipment

Computer hardware is variable and constantly changing. There is a constant drive to make computer hardware faster, cheaper, and store more data. As we saw in the last module, computer equipment (both hardware and software) is very complex and requires many changes over time. The following pieces of equipment, software and protocols are basic to any GII:

- Servers
- Hard drives
- Monitors
- Routers
- Firewalls
- Network Software
- Network Hardware
- Security, Authentication, and Authorisation Services
- Security and Encryption

- Portals

Over time, all of these components will wear out or become obsolete, and will have to be replaced. Obsolescence may occur for technical, or for financial reasons. New equipment may allow for a dramatic decrease in operational costs, and even though it is not technically obsolete, this equipment may be replaced by newer, less expensive equipment.

Data Updating Policies

Several issues must be resolved in data updating, including data availability, accessibility and applicability (Williamson *et al.*, 2003). The updating of GII data requires policies that support decision structuring tools emphasising information search and value, e.g. network analyses, decision tree models and confidence rating schemes for communication. Tools must be available to help decision makers design and circulate alternatives for strategic planning. These may include smart databases that codify, catalogue and enable access to agency decisions, with analyses of such decisions helping to improve future decision-making. Other tools are required to quantify and visualize the accuracy assessments of GII data.

Issues surrounding data accessibility require tools such as guides to data experts who can explain database contents, online links to knowledge bases, enabling problems to be solved, and publicly-accessible technology tool boxes, which are interoperable and enable the design of decision support systems.

There are a number of tools to help assess data applicability. Problem visualization aids can be used to display data and graphical symbols for the representation of changing problems. Tools which display uncertainty representation schemes allow the assessment of uncertainty and risk in data and models. Trade-off analysis models allow users and managers to understand what is being lost or gained in the selection of one option over another. Finally, chat utilities and bulletin boards allow analysis teams and decision makers to exchange views on changing analyses and decisions.

Data must be available, accessible, and applicable across networks in a GII. Rapidly changing technology, competition in software and hardware, and variable priorities in data and information production may cause network incompatibilities. Accordingly, special attention must be paid to software and hardware compatibility, and common languages or translation services. It is important to monitor changes to hardware, software, human management and project priorities across networks, locally, nationally and internationally to ensure that a GII continues to be functional and relevant.

Metadata Creation

Metadata, which is data that describes other data, enables people to understand the software and information networks basic to GII development. An item of metadata may describe an individual datum, or content item, or a collection of data including multiple content items. The metadata required for effective data management varies with the data and its use. Metadata may be used to simplify the search for resources on computer networks, and provide additional information to users of the data it describes. It also helps to bridge the semantic gap, by telling a computer how data items are related and how these relations can be evaluated automatically. It is therefore possible to process even more complex filter and search operations. Metadata has become important on the World Wide Web because of the need to find useful information from the mass of information available. This is particularly important for GII development, and information interfaces between the web and geospatial databases.

4.2.4 Lessons Learned

In this module, we have learned about GII development, administration and maintenance. These have been described in some detail. The following points are especially important to the understanding of this topic.

- GII development is an ongoing process, in which the second generation of GII developers learned from the first, and both groups are increasingly aware of the broader social, economic and technological issues surrounding GII.
- Different agencies at different scales have important roles to play in GII development, and the leading or subsidiary role does not imply importance; all are indispensable to GII development.
- GII development is a very complex process, involving continual learning and relearning, testing and strategic planning. In this sense it is both a technological and a social process.
- The socioeconomic justification for GII development is important from a political perspective, as it gives governments a rationale for the allocation of funds. It also demonstrates the utility of this method.

4.3 Conclusions

In this module, we examined the architecture of GIIs. Using the hardware, network, and software components that we described in the previous module, we examined how these could be combined to create a system that was greater than the sum of its parts. The key to this synergism is the development of services, which are standalone applications running on servers that can be chained together in a process called pipelining, to produce virtually any result.

We examined a model of how the Lithuanian GII might work, and the components into which it was divided. The Discussion Model was composed of five subsystems: the Security Subsystem, the User Management Subsystem, the Data Warehousing and Formatting Subsystem, the Data Preparation Subsystem, and the Financial Subsystem. Each of these subsystems consists of one or more servers running one or more services which are connected together with the network, and which are carefully insulated from the Internet by a series of Firewalls and Proxy Servers. The use of a Reverse Proxy Server means that there is only one entry point to the GII from the Internet, that all requests are carefully filtered to ensure that they are valid, and that all requests that go beyond the final Firewall are routed to a server whose name is not visible to the public. These security measures ensure that the network behind the Firewall can be kept accessible to the staff that contributes to the GII. It also means that computers need not be duplicated, because the source data for the GII resides in directories on the servers of the ministries that contribute to the GII.

The Discussion Model described in this module contains all of the key elements of GII. The system described gives users the ability to search and examine metadata, to identify appropriate data, and to purchase data of interest on the Internet. Users can automatically create and view standard data products without human intervention, or pay to have a cartographer and/or scientist create a non-standard data product for electronic delivery. Data can be delivered over the Internet, and individuals without Internet access can identify and order data products for delivery as paper maps or as digital data on optical media such as a CDs or DVDs.

Many other possible services could be added to the Discussion Model, including services for mobile users such as Location-Based Services, Geocoding Services, Navigation Services, Directory Services, and services for scientific data collection such as a Sensor Web. Once the Lithuanian GII has been established, the system will become increasingly powerful as new technologies become available to exploit the rich data environment

This module has examined the development and maintenance of GII. It is evident that all the topics examined in the previous modules are relevant to this final topic. GII development has been revealed to be a complex technological, social, economic and managerial process. In this sense, GII is inherently geographical, with spatial variations across the world and encompassing methods of natural, social and technological sciences. These points are particularly important to countries undergoing GII development, as the shared experiences of other countries, the unique local context, and the shared political, economic, social and environmental contexts have been shown to be equally important. These points are important at the international, national, local and institutional levels. Considering the topics covered in the earlier modules, it may be seen that these topics are also very important at the individual user's level, and this therefore bridges the gap between the policy issues and individual, private GI users. The evidence in these modules is that a GII cannot function without people, procedures, and organizations. People's cooperation is interlinked, transcending national borders, public and private organizations, cultural perspectives and training systems. The procedures adopted were also important. Although certain methods were described as universally beneficial, one of the most important issues discussed was flexibility and context, hence the global choice of examples. These procedures were the background to the design and maintenance of the GII, and illustrated the fundamental fact that organizations, their mode of creation, design and maintenance is vital to the effective functioning of GII.

Module Self-Study Questions

1. Given the recent attacks on the Estonian computing infrastructure, do you consider that the Discussion Model presented in this lecture features adequate security? What trade-offs are there between security and usability of the GII? At what point does security becomes so onerous that it prevents the Lithuanian GII from meeting the needs of the people?
2. What are the advantages of building the Lithuanian GII at this point in time? What are the disadvantages?
3. The Discussion Model attempts to meet the needs of many users. To establish such a system will require a large amount of resources, both human and financial. Can you suggest a way to divide up the Discussion Model into 3 or 4 subunits that could be constructed more easily using a phased approach?
4. Compare the main issues in the first and second generation of GII developers. What do you think are the main issues that influenced the differences between the two generations?
5. Give some arguments justifying public expenditure on GII development. What are some possible social benefits of GII development?
6. What are the main functions of lead and subsidiary agencies in GII development? Give examples of these two types of organizations, and describe how they may facilitate GII development with reference to specific types of projects.
7. In the testing of GII software, describe the benefits of Alpha, Beta testing and Going Live. What would be the main problems?

Suggested Readings

- Creating and Publishing Metadata in Support of Geospatial One-Stop and the NSDI (http://www.geodata.gov/gos/metadata/CreatePublishMetadata.pdf, Nov. 28, 2007)
- Groot, Richard & McLaughlin, John (2000). *Geospatial Data Infrastructure: Concepts, Cases, and Good Practise.* Oxford: Oxford University Press.
- Nebert, Douglas D., Editor (2004). Developing Spatial Data Infrastructures: The SDI Cookbook, Version 2.0 25 January 2004. Global Spatial Data Infrastructure
- American Library Association, *Task Force on Metadata Summary Report*., June 1999 (http://www.libraries.psu.edu/tas/jca/ccda/tf-meta3.html, Dec. 28, 2007)
- Groot, R. & McLaughlin, J. (2000). *Geospatial Data Infrastructure*. Oxford University Press, Oxford.
- Masser, I. (2005). *GIS Worlds: Creating Spatial Data Infrastructures*. ESRI Press, Redlands.
- Williamson, I, Rajabifard A. and Feeney, M.F., Eds. (2003). *Developing Spatial Data Infrastructures: From Concept to Reality*. CRC Press, Boca Raton, FL.

Assignments

Assignment 5: GII Development, Administration, and Maintenance

References

- Creating and Publishing Metadata in Support of Geospatial One-Stop and the NSDI (http://www.geodata.gov/gos/metadata/CreatePublishMetadata.pdf, Nov. 28, 2007)
- ESRI, 2007a *Arc IMS notes* (http://www.esri.com/software/arcgis/arcims/about/overview.html, December 13, 2007)
- ESRI (2007b). *Introducing ESRI Image Server* (http://www.esri.com/news/arcnews/fall05articles/introducing-image-server.html , Dec 13, 2007)
- Groot, Richard & McLaughlin, John (2000). *Geospatial Data Infrastructure: Concepts, Cases, and Good Practise.* Oxford: Oxford University Press.
- Lake, Ron, David S. Burggraf, Milan Trninić, & Laurie Rae (2004). *GML Geography Mark-Up Language: Foundation for the Geo-Web.* Wiley.
- Nebert, Douglas D., Editor (2004). Developing Spatial Data Infrastructures: The SDI Cookbook, Version 2.0 25 January 2004. Global Spatial Data Infrastructure
- Nguyen, Anh-Duy (2004). Securing Web Application is through a Secure Reverse Proxy. *InformIT* (http://www.informit.com/articles/article.aspx?p=169534&rl=1, Dec. 5, 2007)
- Open Geospatial Consortium (2007). OpenGIS Catalogue Service Implementation Specification (http://www.opengeospatial.org/standards/cat#overview , Dec 14, 2007)
- American Library Association, Task Force on Metadata Summary Report., June 1999 (http://www.libraries.psu.edu/tas/jca/ccda/tf-meta3.html, Dec. 28, 2007)
- Groot, R. & McLaughlin, J. (2000). Geospatial Data Infrastructure. Oxford University Press, Oxford.
- Masser, I. (2005). GIS Worlds: Creating Spatial Data Infrastructures. ESRI Press, Redlands.
- Williamson, I, Rajabifard A. and Feeney, M.F., Eds. (2003). Developing Spatial Data Infrastructures: From Concept to Reality. CRC Press, Boca Raton, FL.

Terms Used

- Alpha Testing
- Authentication Services
- Authorisation Surfaces
- Beta Testing
- Commercial Off The Shelf (COTS)
- Cost Recovery Model
- Decision Tree Model
- First Generation of GII
- Going Live
- Human Resources Strategy
- Integration Testing
- Lead Agency
- Metadata
- Monetary Compensation
- Non-Monetary Compensation
- Phased Implementation Strategy
- Public Funding
- Second Generation of GII
- Security Services
- Subsidiary Agency
- System Testing
- Trade-off Analysis
- Translation Services
- Unit Testing
- ANSI Z39.50 Metadata Standard
- ArcXML
- Attribute Data
- CEN Dublin Core Metadata Standard
- Clareon
- Clearinghouse
- Coater/Decoder (codec)
- Commercial Off The Shelf (COTS)
- Continuous Data
- DBMS Server
- Data Preparation Subsystem
- Data Warehousing and Formatting Subsystem
- Database Management System (DBMS)
- Default-Deny Firewalls
- Directory Services
- Discovery Metadata
- Discrete Data
- E-Wallet Server
- Exploitation Metadata
- Exploration Metadata
- Extensible Markup Language (XML)

- Federal Geographic Data Committee (FGDC) Metadata Standard
- File Transfer Protocol (FTP)
- Financial Subsystem
- Firewall
- Geocoding Services
- Geography Markup Language (GML)
- Geospatial Data Catalogue
- Geospatial One-Stop Portal
- GIS Data Server
- Global Metadata
- Hypertext Markup Language (HTML)
- ISO 15046-15 Metadata Standard
- ISO 19115 Metadata Protocol
- ISO 23950 Search and Retrieve Protocol
- Image Server
- Load Balancing Server
- Local Metadata
- Location-Based Services
- Map Server
- Navigation Services
- Open Geospatial data Consortium (OGC)
- Ordering Module
- Pipelining
- Proxy Server
- Raster Data
- Reverse Proxy Server (RPS)
- Security Subsystem
- Sensor Web
- Spatial Data Directory
- User Management Subsystem
- Vector Data
- Web Accessible Folders (WAF)
- Web Coverage Specification (WCS)
- Web Feature Service (WFS)
- Web Map Server (WMS)
- World Wide Web Consortium (W3C)
- XML Transformation Language (XSLT)