EXTENDING GEOINFORMATION SERVICES: A VIRTUAL ARCHITECTURE FOR SPATIAL DATA INFRASTRUCTURES

Javier MORALES, M. Mostafa RADWAN

International Institute for Geo-Information Science and Earth Observation (ITC)
P.O. Box 6, 7500 AA, Enschede, the Netherlands
(jmorales, radwan)@itc.nl

Commission IV, WG IV/4

KEY WORDS: **Abstraction, Design, Management, Modelling, Services, Spatial Infrastructures, Systems, Requirements**

ABSTRACT:

In the early 90's research on the concepts of spatial data infrastructures focused on ways to improve the capabilities of organizations to supply, share and discover the fundamental data required in their businesses. As a consequence geographic data has become more broadly available, however, continuous change in user requirements driven by factors like new technology, competition, etc., determines that customer satisfaction depends not only on the availability of data, but also in its suitability to the users' particular application domains. Through the integration of data, processes, operations and applications into an infrastructure capable of generating more specialised information services, highly suitable value added data can be obtained that suits specific user communities. Designing such geo-information systems, requires of an appropriate design methodology that supports abstraction, modularity and other mechanisms to capture the essence of the system and help controlling complexity. Numerous concepts to address this idea like system modelling, virtual enterprise, etc., have emerged in the last years. Such concepts properly applied to the geospatial information industry can take us beyond the SDI as just a mechanism for accessing geospatial data to include also demand oriented finished products and services. In this paper we describe an ongoing research project that aims at developing a methodology that enables the modelling of geo-information systems in an integrated perspective. Such methodology will help to develop the SDI as a set of interconnected companies that are integrated (virtual enterprise) and are sharing artefacts (data, value-added products, services) that are located along their distributed nodes and have an economic value. These companies are assembled to operate in a collaborative work perspective, to achieve business goals. The methodology aims at supporting the design process of the system, and also helps enforcing that geo-information systems are reliable, maintainable and compliant with the changing requirements.

1. INTRODUCTION

The goal of durable operationalization of Geo-information Science and Earth Observation will depend increasingly on evolving Spatial Data Infrastructures (SDI) and effective and efficient organizations (for example National Mapping Agencies and National Cadastres) that provide broadly needed fundamental data for a multiplicity of GIS application domains.

In the early 90's research has been focus on the concepts of Spatial Data Infrastructures (SDI) and on processes to improve the performance of organizations that supply the fundamental data in the SDI. Results of such research in SDI can be seen in (Groot R., et all, 2000). Meanwhile new developments in the industrial and service sectors are centring on the concept of Virtual Enterprises (VE). The objectives of the Virtual Enterprise concept are better customer satisfaction, reduced time-to-market and adaptation to changes in the continuous change of requirements. They are called "Virtual" because of their temporary nature, seizing certain, often short-lived business opportunities exploiting the opportunities offered by the ICT. VE is composed of functions provided by the participating organizations (enterprises) and structured and managed in such way that they present themselves to third parties as one enterprise. Applied to the geospatial information industry such concepts take us beyond the SDI as just a mechanism for accessing geospatial data to include also demand oriented finished information services. In turn this extended concept confronts us with the necessity to critically review how the conventional missions of the traditional Surveys will be affected and carried out by organizations that will operate largely beyond their traditional organizational borders. Good examples of the implications of these new concepts are "The National Map" (U.S. Geological Survey, 2001) and the "one-stop-shops" on the Internet to provide integrated public services.

Hence methodologies from manufacturing and service industry were studied to appraise their applicability to the new geoinformation production and dissemination environment. The research activity needs to continue along these lines to develop SDI as a set of interconnected companies that are integrated (virtual enterprise) and are sharing artefacts (data, value-added products, services) that are located along their distributed nodes and have an economic value to provide specialized services, this taking into account reference models and standards that provide a framework for specifying open, flexible and distributed systems.

Such integrated reference models will provide the focal point, around which the business operations in the SDI are designed, implemented, managed and improved and/or new business opportunities are identified. It will also support the assessment of SDI performance in 'totality' along the various operational dimensions, i.e. quality, time of delivery, cost, optimum use of resources, monitoring of changes in the surrounding environment as well as the capability to adapt at both the institutional, the organization's business and operational levels

Our starting points to develop methodologies to model the SDI as an integrated enterprise to achieve well-defined business goal focus on investigating:

- How can SDI (Spatial Data Infrastructure) operate as a set of interconnected companies that are integrated (virtual enterprise) and are sharing artefacts (data, value-added products, services) that are located along their distributed nodes and have an economic value? How can these companies be assembled to operate in a collaborative work perspective, to achieve business goals?
- How can the existing reference models and standards that are developed for open and distributed systems be used to describe and model various interactions (such as goals, processes, information, resources, rules as well as the surrounding changes) in such a SDI?
- How can the SDI integrated enterprise model be validated and used to implement requirements and to assess the SDI performance?

These broad domains will provide the context for the development of the "enterprise model" for a SDI. The result will be a mechanism for capturing knowledge of a SDI system in terms of simple components that can be assembled into large functional specifications. Our research concentrates on the definition, development and validation of the underlying principles that support the concept.

2. GEO-INFORMATION SYSTEMS

The change of emphasis in geo-information production from just data towards diverse products and services forces a change in the design of the system that handles such production. Diverse products and services imply that requirements are dynamic.

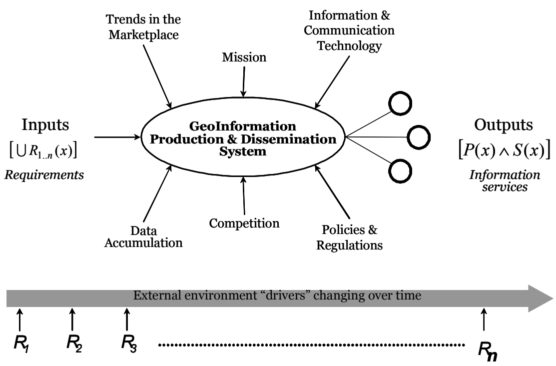

Figure 1. GIP environment & external drivers

To be effective in such conditions, one would prefer to have an adaptable system rather that having to re-engineer it every time. Figure 1 shows the environment surrounding geo-information provision systems. External factors like technology, competition, policies, etc. are the forces that drive the requirements and define the responsibilities (functionality) of the system at any point in time. Continuous change of requirements $R_{\{1..n\}}$ through time affects the system and subsequently impacts its behaviour and responsiveness. Possible outputs in terms of Products *[P]* or Services *[S]* are defined by the inputs *[R]* to the system. Inputs varying over time mean that the system has to adapt continuously in order to respond satisfactorily. We believe that if geo-information systems are to remain accepted as part of large scale business applications, the use of suitable design methodologies is crucial to attack this challenge.

By studying existing geo-information systems, we observe that most of them were design using data oriented design methods and some few using some basic forms of object concepts. They operate with an architecture suitable to acquire and store geographic data that can be retrieved at will. This architecture facilitates the production of certain maps but few diverse products and services. An interesting aspect however, is the existence of a mechanism for providing access to data produced by these systems. This data sharing mechanism is known as the Spatial Data Infrastructure (SDI). A SDI is defined as: the relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data. The GDI provides a basis for spatial data discovery, evaluation, and application for users and providers within all levels of government, the commercial sector, the non-profit sector, academia and by citizens in general (Douglas N.D., 2001).

We want to take advantage of the existence, understanding and maturity of this concept for data sharing and build upon it a system concept for the generation of a wider range of quality value-added information services as defined by the market. For this purpose we introduce the term GSI (Geo-information Service Infrastructure). A GSI is a system where specialised information products and services can be obtained by exploiting an infrastructure of interconnected data nodes (data repositories), data brokers, service providers, service brokers and clients. The benefits obtained from the GSI are based on a set of artefacts located along the distributed nodes which have an economic value; these artefacts can be assembled to perform operations within the infrastructure, resulting in an specialised artefact that has a value equal or larger than the value of the artefacts used. This behaviour can be regarded as a ``value-added discrete system". A service is created by integrating these artefacts and generating a functionality that satisfy a particular set of requirements. This flexible design based on connectable components is an appealing way of looking at geo-information provision systems but it imposes a great challenge from the design point of view. We dedicate the following sections of this paper to present a design methodology to support the development of sufficiently reliable, maintainable and conformant GSI systems.

3. A DEVELOPMENT METHODOLOGY

A very important consideration in the conceptual modelling of information systems is that models have to describe structural and behavioural aspects of a system (Graham I., 2001). Traditionally, the modelling of both aspects in geo-information systems has been treated separately, and in reality little or none focus has been put to the behavioural part of such systems. Other information based domains have already incorporated into their design methodologies techniques that integrate both aspects. Numerous techniques for this type of conceptual modelling have emerged in the last years like UML (OMG, 1999), IDEF (Mayer R. J., et all, 1995.), etc. Researches at the University of Twente (the Netherlands) have worked for the past several years in the creation of a design methodology (AMBER) to support the development of various types of system models. This method was created to support effective design of distributed systems (Ferreira Pires L., 1994; Quartel D., 1998). We have taken behavioural modelling ideas (AMBER), together with concepts from abstract data types and specifications of systems, to define a methodology devoted to modelling both structural and behavioural aspects of geoinformation systems and more specifically GSI systems.

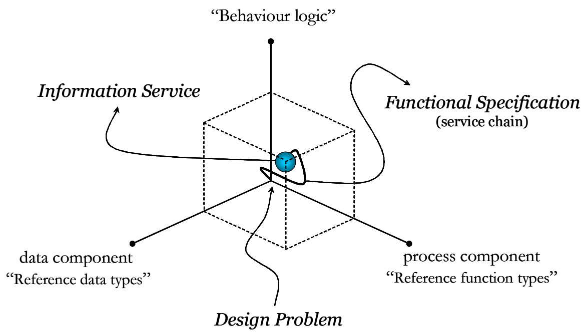

Figure 2. Modelling dimensions

The GSI system's principle of logical distribution provides a general, flexible and evolutionary approach. It offers advantages because it forces one to think in terms of objects, modularity, flexibility, structure and responsibility or, in other words, architecture. However, architecture implies the identification of components, independency, atomicity and, furthermore, interaction, relation and synchronisation between independent components. Such consideration forces one to look at those systems from different perspectives, and at different levels of detail (D'Souza D.F., et all, 1999). We have identified three views to consider for a proper description of the different aspects of a GSI system. Every one of the three axes of the diagram in figure 2 represents an aspect of the system. The data axis represents the information that is used, manipulated and/or generated by the system. The design concepts defined along this axis are used to capture information about *data components* of the system. Data components can be describe in terms of their structure, relationships and associations; the process axis represents the geo-processing capabilities of the system. Design concepts along this axis are used to describe *process components*. Process components are used to generate or modified data components. Process components can be applied to the data components but are independent of data objects themselves; the control axis represents design concepts that allow the definition of *service chains*. This includes the conditions and constraints that define the interactions between data and processes components. The origin of the tri-axes reference system represents the starting point of a processing flow that is activated by a design problem. Design problems are the various inputs (request) to the system that have to be converted into products and or services. The path taken through the cube is a functional specification of a service chain that satisfies a design problem by making use of a proper combination of artefacts. The execution of a predefined service chain generates a desired product or service.

In section 2 we explained the basic GSI concept. From that definition we identified the composing artefacts of a GSI system and they can be categorised as follows:

- Data, stored as datasets with proper documentation (metadata),
- Operations to perform transformations on datasets,
- Processes to generate non existing datasets,

These groups of artefacts constitute the foundation of the system (Figure 5). Existing datasets stored in data nodes, and processes and operations provided by the service nodes, have to be described properly within the corresponding view before they can be integrated into a service chain. One additional feature to consider here is the potential evolution in terms of the number of elements that the system may include. To address the above mentioned characteristics our structuring approach focuses in three perspectives: first the *external perspective* captures the responsibility of the system that is defined by the relations with the surrounding environment; next the system's constituent or forming elements (components) are described in the *internal perspective*, and then the possible interactions between the forming elements to achieve the various functionality are modelled in the *distributed perspective*.

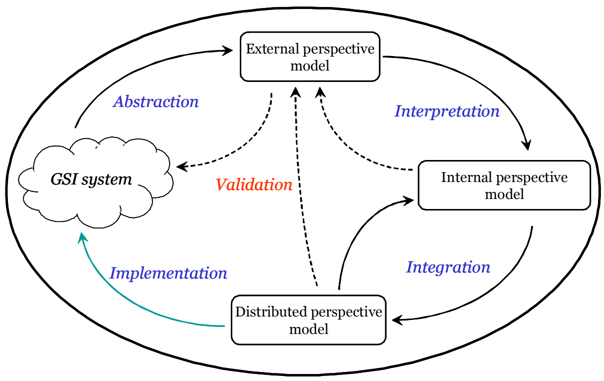

Figure 3. Modelling steps

We have defined three different levels of abstraction for the manipulation of GSI systems the external, the internal and the distributed (figure 3). The external perspective aims at identifying and explicitly delimits the scope of the system under development and that helps to determine the objectives of the development process. This definition of scope and objectives needs initially not be as detailed as a complete service description of a system. This step is done such that further more detail specifications can be guided by these definitions. Such scope can be defined for example as short statements describing the product(s) and/or service(s) to be delivered to the system's external users. The internal perspective focuses on the formal definition of the elements of the GSI system that fall within the selected scope and can satisfy the determined objectives. The output of this face is a set of system components of two types: the data-objects and process-objects; data-objects are modelled using data types which characterise their purpose; process-objects are modelled as complex actions depicting the data that they generate or require and the service that they perform. The distributed perspective aims at describing the internal system

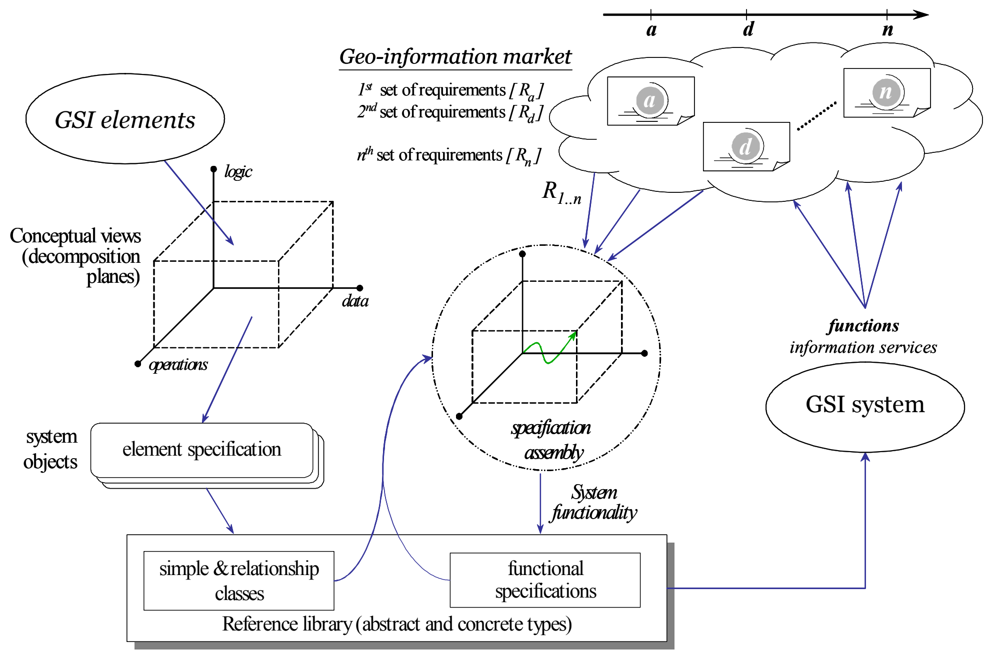

Figure 4. Development Methodology

structure in terms of composition of various components that interact to perform the system functions. This is regarded as system behaviour and shows how groups of interacting components work together to achieve an observable functionality of the system. Many different compositions are allowed to be define that may implement the same functionality. Or even functions that look the same but the quality of the generated services may vary.

To acquire the necessary information for the generation of the various system models it is crucial to follow a structuring method. Figure 4 depicts a structuring notion for GSI system; the top-left part of the figure shows how the system has to be analysed as a collection of elements that are in principle independent form each other. Using categorisation, those elements are arrange according to their correspondence to the three system views (figure 2). Every contributing element has to be described to enable its participation in different system functionality. For that purpose a formal definition of the elements is produced and stored in a reference library. At this point each element has its own definition and can now participate in different service chains that will deliver services to the clients. To define these service chains we used the responsibilities or requirements that are being continuously identified from the geo-information market (top-right of the figure) and that have to be supported by the underlying system. These responsibilities (sets of requirements) are used as design problems. Suitable interactions of components that are compliant to the different design problems are generated. This step is called specification assembly. During this step a path through the cube is defined it represents a behaviour definition (specification assembly). Every behaviour definition will be a sequential arrangement represented by the relations between the participating elements, including conditions and alternatives. The logic that drives these relations is based on pre and post conditions that are evaluated to identify the correct sequences. These functional specifications suitable to deliver particular services are stored in the reference library.

Functional specifications can be reused in combination with additional elements to define specifications to define more complex functions. All this stored specifications can be enacted with a workflow management system or a similar tool to physically generate a desired service.

4. GSI ARCHITECTURE

The GSI is defined as a set of interconnected companies (GSI Nodes, Figure 5) that are integrated and sharing artifacts (business goals, strategies, data, processes, value-added products, systems, etc.) that are located along their distributed nodes and that have an economic value, to provide specialised services, all this within a framework of reference models and standards that provide the mechanism for specifying open, flexible and compliant systems.

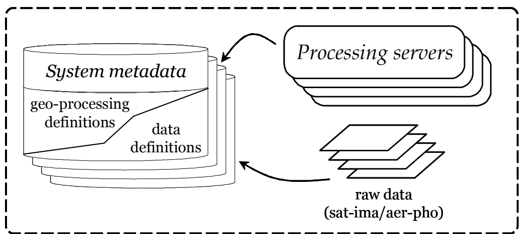

Figure 5. GSI Node

The enterprise model for GSI consists of various models (each can be broken into more detailed sub-models), such as:

- Goals/Business Vision: describes the purpose of the business, what to achieve, the goal structure, and illustrates problems that must be solved in order to reach those goals;
- Production Model: represents the geometric and non-geometric features as well as design details of products, their logistics and life cycle;

- Process models: indicate the set of operations (or actions) to be performed to execute the enterprise activities and do the work. They model the functionality and behaviour of the enterprise in terms of processes, activities, basic functional operations, sequences in which things must be done and identifying triggering events;
- Information Models: describe information entities and their relationships, information semantics and the federation mechanism of, mostly heterogeneous, databases;
- Engineering Models: describe the distributed system infrastructure;
- Resource Models: describe the technology aspects and specify the characteristics, capacity, management policies and possible actions of H/S resources and their configuration to perform enterprise activities;
- Organization models: document the organizational structure in terms of decision levels, organizational units, hierarchy;
- Economic Models: provide a cost-oriented analytical view of the enterprise to evaluate the cost-effectiveness of the various parts of the enterprise, the pricing structure, etc;
- Optimization and Decision Making Models: deal with issues from operations research and control theory and used by the DSS;
- Rules can be functional, behavioural and structural, constraining some aspects of the business such as processes and resources. They also include the rules enforced by law and regulations.

Figure 5 shows the high level architecture of a GSI Node. A GSI node is a participating entity in the infrastructure. It provides either raw data or geo-processing functions. These two types of elements are described as data-elements or process-elements. List 1 shows the specification of a field data-element. It represents a data type that characterizes parcel data, and it is available at a GSI Node and can be incorporated into a service specification. The description contains other data types used by this data-element like crop, and it also includes attributes, constraints and operations that fully defined the data-element type.

data element **Field**
  components
    crop(ID : *int*) : Crop → *theme(regions)*
  attributes
    area : *real*
    access_tax : |AccessTax|
    rainfall_categ : *int*
    temp_categ : *int*
    num_cr : *int*
    location : *geobj(regions)*
  operations
    addcrop (ID : *int*)
  constraints
    crop(id).area +...+ crop(id+num_cr).area ≤ area
end;

List 1. Field data-element

Fig * Shows the definition of a process-element. In this case it is called generatingAOI. This description is known as a behaviour block, it presents the action executed to carry out with the process and it also depicts the relationships among those actions (Ferreira Pires L., 1994; Quartel D., 1998).

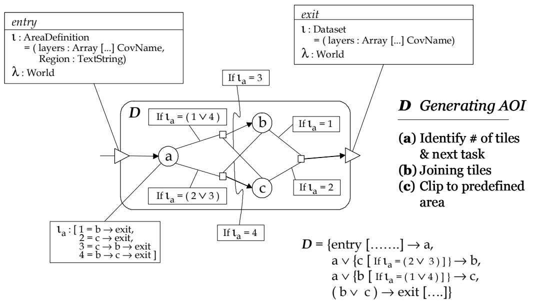

Figure 6. generatingAOI process-element

service element **TIC-service**
  attributes
    …
  data elements
    taxform : (ID : *int*) : set(TaxForm)
    …
  process elements
    generatingAOI (…) {*A*}
    cropClassification (…) {*B*}
    taxfactorsdefinition (…) {*C*}
    taxcalculation (…) {*D*}
  behaviour
    {√ → entry
    entry […] → A[if ι_{entry}.Location ∩ Enschede ≠ null],
    $A \vee \{ C$ [if $|\tau_B - \tau_C| > 5$] } → B (λ : all provinces)
      [λ_B = ιentry .Location],
    $(B \wedge \neg D) \rightarrow C$ (λ : all provinces) [$\lambda_C = \lambda_B$],
    $(B \wedge \neg C) \vee (B \wedge C$ [if $|\tau_B - \tau_C| \leq 5$) → D
    D → exit […] }
end;

List 2. TIC-service definition

A service definition is achieved by combining elements definitions. List 2 shows such a specification. In this case the specification of the TIC-service (Tax identification and calculation service) consists of four parts: attributes, data-elements, process-elements and behaviour. For the first part of the example we do not consider any user defined attributes but all of those required to qualify the service are to be included in this part of the specification. The following part contains all the data elements from which instances of objects are required for use within this service definition, in this example we defined a set of objects of the type TaxForm. In the same way the process element part includes the necessary activities that implement the service. These activities contain internal actions that were defined within every individual process element in their correspondent behaviour block. To simplify the behaviour part of the example we have renamed the activities as letters, *generationAOI* is {*A*}, *cropClassification* is {*B*}, etc. The last part of the TIC-service specification defines all the possible interactions between the participating elements as a behaviour definition. This definition starts with the process-element *generatingAOI* that determines the areas assigned to a *TaxForm* that require tax calculation; next, if the stated constraints are satisfied, the process-element *cropClassification* is perform to identify existing crops within the defined areas; The remaining steps are concerned with the tax calculations of the various fields. An important aspect to notice here is that these process-elements are in principle independent of each other, they may exist within one organization or they may be supply by an external service provider. The importance is that they are specified in a way such that they can be used in the specification in spite of their physical location or specific implementation. Another advantage of the behaviour definitions is that those definitions do not only specify the interactions between the elements that form the service, but they can be used to automate the service by means of a workflow management tool.

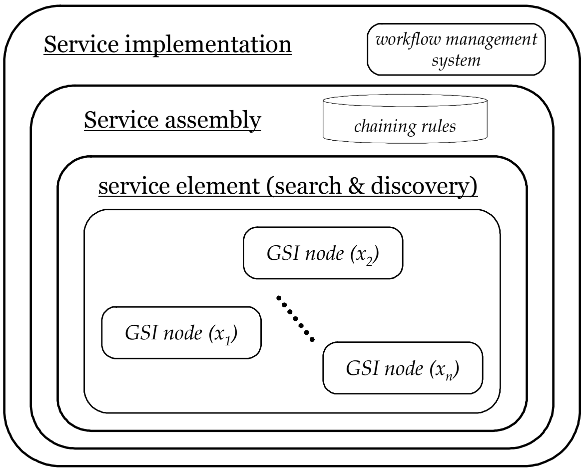

Figure 7. GSI architecture

Figure 7 presents the architecture upon which the GSI lies to provide the desired level of services. The inner layer consists of all the interconnected nodes containing functionality (functions or data) to be share among all participants. These nodes contain metadata descriptions at the system level (List 1, figure 6). The following layer provides functionality for search and discovery of elements to generate service definitions. This search is performed to identify elements that satisfy a required processing step in a service chain definition. The third layer (service assembly) provides functionality to logically connect the various elements into a coherent specification (List 2) that is compliant with an information service request. The outer layer provides functionality to execute process definitions based on the design made in the previous layer.

5. CONCLUSIONS

In this paper we presented the current status of an on-going research that is being carried out in cooperation between the University of Twente, the Netherlands, specifically the Architecture of Distributed Systems group, and the International Institute for Geo-Information Science and Earth Observation (ITC), the Netherlands. Our aim in this research is to define a methodology for the generation of high level generic specifications of geo-information systems components that facilitate their integration for the generation of services that are more compliant with the ever changing requirements of the geo-information users.

The GSI is defined as a set of interconnected companies that are integrated and sharing artifacts (business goals, strategies, data, processes, value-added products, systems, etc.) that are located along their distributed nodes and that have an economic value, to provide specialised services, all this within a framework of reference models and standards that provide the mechanism for specifying open, flexible and compliant systems. During the development of this research project we have recognised the need for a proper mechanism to define, manage and publish metadata at the system level. Since the vision of future systems not only the GSI, is the integration of a broad range of functionality into services to users, it is imperative to be able to express services and functionality in a formal and generic manner. In the field of geographic information large research projects have been executed to define the standards and methods for metadata management at the data level. In this document we have presented our attempt to describe system functionality at the metamodel level. A procedure for the definition and assembly of elements was shown, and concepts for the specification of both behavioural and informational aspects of geo-information systems were addressed, issues that in our application domain has been treated separately and more over no attention had been placed to the behavioural aspect, aspect that becomes very relevant when it comes to share functionality in an open system. We are sure that in the next stages of this work much more detail and valuable information will be develop with respect to these subjects.
Tool support is a crucial element for the success of large scale development projects like the GSI. We do not yet have sufficient automated tools to help in the validation of models, but what is more important even are the tools the manipulation, change and maintenance of the specification models. This is a very relevant issue, because once a particular service functionality specification that is shared in a repository is changed, the effect should be reflected to the services build on top of that specification, and without proper management of this aspect, inconsistencies will rapidly develop and the whole architectural concept will become uncontrollable.

REFERENCES:

D'Souza D.F., and Wills A. C., 1999. *Objects, Components and Frameworks with UML*: The Catalysis Approach. TheAddison-Wesley object technology series. Addison-Wesley, Reading, Massachusetts.

Douglas N.D., 2001. Developing spatial data infrastructures: The SDI cookbook. Technical report, Global Spatial Data Infrastructure Organization.

Ferreira Pires L., 1994. *Architectural Notes: a Framework for Distributed Systems Development*. PhD thesis, University of Twente, Enschede, The Netherlands, Ph.D. Thesis no. 94-01.

Graham I., 2001. *Object-Oriented Methods: Principles and Practice*. The Addison-Wesley object technology series. Addison-Wesley, Reading, Massachusetts, 3rd edition.

Groot R., McLaughlin J., 2000. *Geospatial data infrastructure: Concepts, cases and good practice*. Oxford University Press, New York.

Mayer R.J., Menzel C. P., 1995. Information integration for concurrent engineering (IICE): IDEF3 process description capture method report. Technical report, Knowledge Based Systems Incorporated.

OMG, 1999. Unified modeling language specification. Technical Report Version 1.3, Object Management Group.

Quartel D., 1998. *Action Relations: Basic design concepts for behaviour modelling and refinement*. PhD thesis, University of Twente, Enschede, The Netherlands, Ph.D. Thesis no. 98-18.

U.S. Geological Survey, 2001. The National Map: Topographic mapping for the 21st century. Report, Office of the Associate Director of Geography, Reston, VA, USA.