

Extending the Quality Concept in Geo-Information Processing

Richard Onchaga

International Institute for Geo-Information Science and Earth Observation (ITC)

Abstract

Recent years in geographical information handling have been marked with constant change and innovation. Geographical information systems and enterprizes have continued to gravitate towards distributed models based on open systems technologies. Increasingly, successful geographical information markets are predicated on robust service infrastructures that offer guarantees on performance, availability, reliability and other critical quality characteristics. This makes a broader concept of quality that surpasses classical notions of spatial data quality and that comprises dynamic aspects of distributed geoprocessing in an increasingly complex Internet environment necessary. The paper reviews the trends shaping the future of geographical information markets, and existing quality frameworks and their limitations in view of current trends. Quality of service is introduced and positioned in the context of distributed geographical information processing as motivation for investigating the scalability of emerging geographical services infrastructures.

Key words: geographical information, performance, quality of service, service, quality

1 Introduction

Geographical information (GI) processing is gravitating towards distributed models based on open systems technologies. This comes in the wake of growing need to share distributed spatial resources, integrate geographical information systems (GIS) with mainstream enterprize information systems and better address the requirements of emerging markets especially in the realm of non-expert GIS application domains including wireless and mobile location based services. The quest in GI system research in recent times has thus been towards evolving standards for specifying interfaces and protocols for software systems that can be deployed in an open environment to provide a wide range of geoprocessing services [4, 1]. Meanwhile, GI providers continue to adopt flexible structures and new business models centered on the virtual enterprize concept to take advantage of advances in business networking and e-commerce solutions [32]. Increasingly, network-based geographical information processing and dissemination is emerging as an inevitable future characteristic of geographical information markets as applications composed of distributed autonomous and interoperable services proliferate and become integrated in enterprize business processes.

Users of distributed GIS services will nonetheless expect levels of performance, availability, reliability, and other important quality characteristics comparable to centralized systems. Providing services with acceptable levels of quality in an Internet environment is a big challenge, which is further exacerbated by the growing complexity of the Internet [25] and the unique character of spatial data. Spatial data are large volume phenomena with complex data structures and are computationally intensive [29, 7, 27]. These characteristics have great impact on the scalability of the distributed computing infrastructure, system and network performance considerably constraining deliverable service quality.

As the technology matures and distributed geoprocessing becomes the norm, the challenge of providing guarantees on the quality of the services delivered increasingly becomes critical. Quality

in the geographical information domain has for long been centered on spatial data and its fitness-for-use [2, 10]. In the distributed environment however, a data-centric definition of quality is limiting and does not encompass the dynamic aspects of online access, processing and dissemination of geographical information and services. Take the example of simple remote data access; a high quality dataset delivered with much delay over a lossy channel and in an incompatible format is not so useful, neither is a poor quality dataset even when delivered with insignificant delay and loss and in compatible format. Clearly, quality in the open environment takes a broader scope beyond classical notions of spatial data quality. In addition to quality of the spatial data being shared, exchanged and processed, issues of service performance, availability, reliability, security, cost and network performance will influence the quality of the service as perceived by the end-user.

The rest of the paper is organized as follows. Section 2 reviews the trends that are shaping the future of geographical information markets. Section 3 outlines the existing quality frameworks highlighting limitations. In Section 4, we introduce the concept of quality of services in the context of emerging geographical services infrastructures and Section 5 provides a summary.

2 Fundamental Trends in GI Environments

The explosive growth of the Internet continues to revolutionize the way modern day business is conducted and services provided. In recent years geographical information systems (GIS) and enterprises have continued to evolve towards distributed models in order to better exploit the potentials presented by the Internet computing paradigm. GIS systems have exhibited sustained evolution from stand alone, data-centric stovepipes to distributed models composed of open interoperable services while GI enterprises continue to pursue flexible models in order to leverage advances in business networking and e-commerce. Meanwhile the spatial data infrastructure (SDI) concept which emerged in the 1980s to advance spatial data sharing by taking advantage of the ubiquity of the Internet and its ease of use, has matured and is evolving into an infrastructure for the delivery of geoprocessing services, the so called geographical services infrastructure (GSI). In the following sections we briefly review each of these developments from a quality perspective. Section 2.1 reviews the trends in geographical information (GI) enterprises. Sections 2.2 and 2.3 outline the concepts of the spatial data infrastructure (SDI) and the geographical service infrastructure (GSI) respectively.

2.1 Evolving GI Enterprises

In a bid to improve performance, organizations charged with the responsibility of collecting, managing and disseminating geo-information have exhibited marked changes in their institutional and strategic set-ups in the past few decades [31, 23, 18]. Customer focus has been the overarching principle underlying pervasive re-engineering programs as providers strive to meet growing demand for a broad range of services and information products. Strategic partnerships and outsourcing arrangements are now commonplace with some providers (e.g. the Dutch Kadaster) expressly providing for user councils in order to pro-actively align corporate strategy with user requirements. Development and deployment of elaborate quality management systems in accordance with ISO standards is another strategy that has been vigorously pursued and widely adopted in many enterprises to enhance the efficiency of business processes and achieve customer satisfaction [13].

Nonetheless, the fast pace and global nature of present day businesses coupled with the prevalence of e-commerce and the push for lean enterprises continue to challenge providers of geographical information services and products. Furthermore, it is increasingly evident that it is beyond the ability of a single organization to meet all the service and information product needs of a rapidly growing user community. This demands a new framework to enable GI enterprises to rapidly

change their structures, enter and leave partnerships quickly, change roles, rapidly design and deliver new products and services in changing business environments.

As a consequence, GI enterprizes have shown increased preference for integrated enterprize models centered on the virtual enterprize (VE) concept to exploit opportunities presented by technological advances, achieve strategic objectives and be effective in competitive global markets. A virtual enterprize comprises of a temporary alliance of globally distributed autonomous enterprizes participating in a common product cycle or service chain and sharing resources, skills and costs with the support of information and communication technologies to better attend market opportunities and fulfill corporate strategy [9]. The virtual enterprize is thus composed of functions provided by participating enterprizes and is structured and managed in such a way that it presents itself to third parties as a homogenous whole.

2.2 The Spatial Data Infrastructure

The spatial data infrastructure (SDI) concept was a pioneering vision of a distributed system dedicated to managing and sharing spatial data among several stake holders at different levels of government through appropriate application of information and communication technologies. This is well embodied in the stated objectives of the National Spatial Data Infrastructure (NSDI) of the U.S.A. The NSDI is mandated to improve spatial data availability, minimize duplication efforts in spatial data acquisition and management, provide public access to distributed spatial data resources, and reduce costs related to geographic information handling (www.fgdc.gov/nsdi/nsdi.html).

The concept of an infrastructure to facilitate spatial data sharing continues to attract growing attention world wide [24]. As it spreads and evolves, the concept has come to mean different things in different contexts and communities depending on motivation, funding and policies for its creation and management [33]. That notwithstanding, the SDI encompasses the policies, technologies, standards, fundamental datasets and human resources necessary for the effective collection, management, access, delivery and utilization of geospatial data at affordable costs [26, 12].

2.3 The Geographical Service Infrastructure

The growing market for specialized and customized services, especially in the realm of non-GIS expert domains that include mobile and wireless location-based services, brings to question the relevance and appropriateness of traditional GIS models. As a result there is growing consensus that new models where GIS functionalities are delivered as autonomous interoperable components are needed [4, 28]. The component based GIS model strongly derives from the web services concept, in which case GIS functionality is delivered as autonomous interoperable services that are accessible through the World Wide Web (WWW) and that can be remotely discovered and combined into customized GIS applications or choreographed to perform complex tasks. This novel GIS model promises applications that are platform independent, that can transparently share spatial resources, and that offer increased potential for favorable cost/performance ratios. Furthermore, traditional GIS systems have for long been developed and managed separately from mainstream enterprize information systems, which inhibits enterprize wide application of GIS technology. The new models will greatly ease the integration of GIS functionality with mainstream enterprize information systems and promote ubiquitous geoprocessing.

GIS applications based on distributed services prompt the concept of a geographical services infrastructure comprising interoperable services that can be combined with each other or with other specialized services (e.g simulation tools) to offer advanced services and execute complex tasks as may be required by some end-user. A service can be defined as a collection of operations accessed through an interface and that allow a user to evoke some desirable behavior.

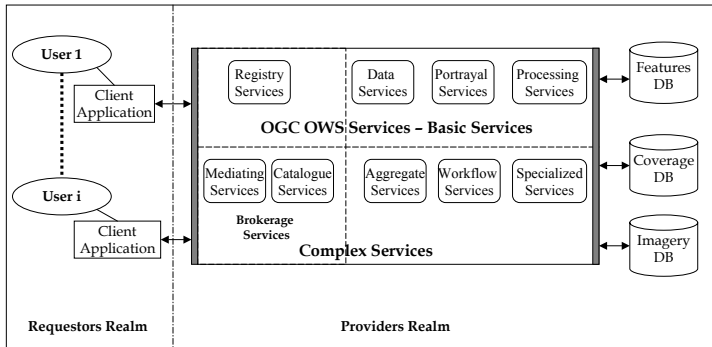


Figure 1: Distributed services

The Open GIS Consortium has defined the OpenGIS web services framework (OSF) that proposes a set of basic services, interfaces and protocols that can be used by any compliant application in a distributed enterprise setting [19] (Figure 1). OpenGIS services are web services that reflect the vision of the Open GIS Consortium (OGC) for spatial data and application interoperability. Ubiquitous deployment of openGIS services is poised to promote establishment of e-commerce architectures by connecting collaborating communities of information providers, maintainers and brokers interworking in shared service chains [19], further advancing the virtual GI enterprise (Figure 2).

3 Quality and Geoprocessing: Existing Frameworks

Quality is a concept that has been widely applied to various aspects of spatial data, its acquisition, processing, management and dissemination, yielding a rich vocabulary of quality-related terminologies and definitions. Quality has for long been focussed on spatial data and its fitness-for-use.

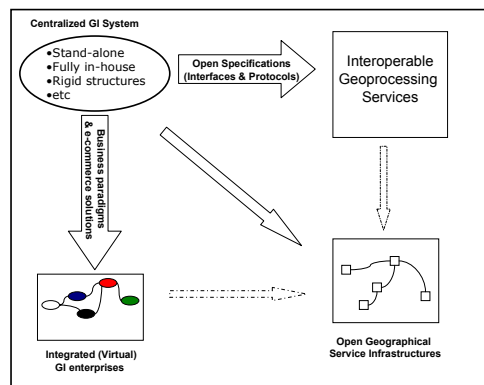


Figure 2: Trends in GIS Systems and GI Enterprises

However, market dynamics and changes in operational environments have seen many GI enterprizes adopt new structures and business models, and increasingly pay greater attention to the dynamic aspects of quality in their business processes. Further, it can be anticipated that changes in enterprize business processes and organizational structure manifest in changing architectures of enterprize systems. We thus envision three closely entwined contexts within which one can define quality viz the contexts of business processes, enterprize systems and spatial data.

3.1 The Context of Business Processes

Recent years have seen considerable change in the geographical information handling and market environments. As a result GI enterprizes have had to rethink their operations, product and service offerings in order to improve performance, stay competitive and expand market niche [31, 18]. Organizational restructuring, business process redesign, deployment of total quality and workflow management systems and quality control procedures (see for example [34, 8, 16]) have thus become commonplace.

The overarching goal is to improve the performance of the business process, hence its products and services. Performance in this context is defined by quality, cost and delay (QCD) [18, 35]. Quality refers to the extent to which the product satisfies user needs also called external quality. The cost of a GI product can be viewed in two ways. First, it can be production cost which is the cost incurred in generation of the product and is directly related to the efficiency of the value chain or business process. It also refers to the price at which the product is made available to the consumer, always considered in the background of perceived quality of the product. Delay is directly related to the responsiveness of the business process and refers to the time it takes to make a product available in the market (time-to-market), and also the time to generate a unit of product. Though the concepts have attracted much interest they are still little developed given the infancy of the markets.

3.2 The Context of Geoprocessing Systems

Enterprizes need flexible systems that can be rapidly adapted to changes in their environment and integrated with other systems when the need arises. Enterprize systems support business processes, store, process, manage and analyze business information in pursuance of strategic objectives. Thus an enterprize system will reflect the strategic goals of the enterprize and its organizational structure.

The trend in recent years has been towards open distributed systems to support the computing needs of collaborating enterprizes dispersed in space. Distributed systems comprise heterogenous computing elements connected via communication networks and offer better cost/performance trade-offs, improved system availability and reliability, and enhanced resource sharing. Interoperability is however a big challenge.

Towards distributed geoprocessing, the OGC has proposed the openGIS services framework (OSF) that defines a set of basic services that can be deployed on the Internet, located and invoked by compliant client applications in a distributed enterprize environment [19]. Basic services can then be combined together to provide aggregated services or integrated with other services to provide more complex and specialized services (Figure 1). It can be anticipated that as the virtual enterprize becomes the norm and distributed geoprocessing gets pervasive, an increasing number of enterprizes will become reliant on distributed infrastructures for mission critical applications.

The ISO¹ and the ITU² have standardized the reference model for open distributed processing (RM-

¹International Standards Organization

²International Telecommunications Union

ODP) to aid the architecting and development of distributed systems and achieve interoperability. The RM-ODP has five viewpoints to enable separation of concerns during design by allowing designers to focus on facets of the design that are of interest. The enterprise viewpoint captures the purpose, scope and policies for the system being designed. The information viewpoint specifies the information the system holds and the processing it carries out. Evidently, spatial data and its quality are among the concerns of the information viewpoint. The computational viewpoint of the RM-ODP supports distribution through functional decomposition of the system into objects which interact at interfaces. Important quality concerns of the computational viewpoint include portability, modifiability and modularity. The core of OGC work has centered on the definition of specifications for interfaces and protocols for geoprocessing computational objects [1]. The engineering and technology viewpoints focus on the mechanisms, functions and technology choices necessary to realize a distributed system. Several quality concerns can be identified for distributed systems (or any system for that matter) and include performance, availability, reliability, security, usability, modifiability, etc [5] and these qualities are chosen to suit the system requirements and its stakeholders.

3.3 The Context of Spatial Data

The classical application of quality concepts in geographical information domains has been to describe aspects of spatial data. For long, data accuracy was the sole concern resulting in elaborate models and techniques to analyze error and its propagation in measurements [17]. With the advent of digital spatial databases and GIS systems, spatial data quality has extended beyond accuracy to include other elements like completeness, consistency, lineage etc. [10].

Growing possibility to share spatial data and apply it for purposes other than those for which it was collected, has led to the general definition of the quality of a spatial dataset as its fitness-for-use in an application context. This definition ties well with that by ISO (9000:2000) that defines quality as the totality of characteristics of a product that bear on its ability to satisfy stated or implied needs [20]. Spatial data quality is currently presented in metadata, which users can employ to determine the appropriateness of a particular dataset for a particular task. Nonetheless, metadata exhibit several limitations [3, 11, 14]:

- Metadata essentially present the providers' view of quality
- Metadata are static
- Metadata are devoid of important information on uncertainties associated with indicated quality and associated risk of using the dataset
- Metadata are very complex for non-expert users
- Semantics and fuzzy objects are little addressed

Several strategies are proposed to address stated limitations. Profiling is proposed to address complexity. Similarly, models that take metadata elements and user requirements as operands to provide an assessment of the fitness-for-use of a dataset and the risk involved in its use are needed [3]. Ontologies for user communities promise to handle issues of semantics [6]. In general, these are nonetheless still very much open issues in research.

4 Quality of Services and Open Geoprocessing

Future geographical information (GI) markets will be dominated by collaborating enterprises interworking and sharing spatial resources in an open technology environment. Pervasive GI processing will exploit the ubiquity of the Internet to promote access to distributed spatial resources and application of geographical information in a wide variety of decision making procedures.

Users in the distributed environment however expect levels of performance, availability and reliability comparable to centralized environments. Performance is about timing. Events occur to which a system must respond, and the time it takes the system to respond to an event generally provides the basic measure of performance [5]. Availability relates to system failure and associated consequences and is characterized by the probability of the system being operational when needed. The reliability of a system is the probability of the systems functioning properly and constantly over a fixed period of time [25]. These quality characteristics have attracted much attention especially with the advent of the Internet and multi media applications. They have generally been studied under the subject of quality of services (QoS) which has been variously defined as:

- QoS is user-perceived performance [15]
- QoS is the collective effect of service performances which determine the degree of satisfaction of a user of the service [21]
- QoS is the degree of conformance of the service delivered to a user by a provider with an agreement between them [30]
- QoS is a set of quality requirements on the collective behavior of one or more objects and may be specified in a contract or measured and reported after an event [22]
- QoS is the ability to provide resource assurance and service differentiation [36]

Central to quality of service are users and their requirements. It should be noted however that user perception of quality is influenced by a wide range of factors, some of which are outside the scope of the application or infrastructure designer. Our focus in geographical service infrastructures will focus on issues that will affect the scalability of the infrastructure. In the context of distributed geographical service infrastructures QoS can be defined as

the level of performance guaranteed over the infrastructure

and will thus be characterized by a collection of values (may be ratio, a maximum, an average etc.) acting on performance related properties of the components that comprise the infrastructure. Support for QoS is necessary failure of which may result in undesirable effects depending on the service and its role within the enterprise. Potential consequences include unpleasant user experience, application failure, loss of business, etc.

As distributed geoprocessing becomes of age and the virtual GI enterprise becomes the norm, the need for QoS guarantees will be heightened even as the applications become larger and more complex. Future GI markets are therefore increasingly predicated on robust distributed infrastructures that can offer guarantees on QoS and other important quality characteristics. Distributed architectures will comprise requestors/consumers and providers of services linked together by brokerage services (Figure 1). Through the use of service metadata, users will discover, access and chain services based on their performance, reliability, availability, conformance to standards, etc. making end-to-end QoS composable, specifiable and predictable. Nonetheless providing QoS guarantees poses considerable challenges. For one, the Internet is an increasingly complex environment and spatial data are complex phenomena with characteristics that can adversely impact performance

and availability in the distributed environment i.e. large volumes, complex data structures and high computational intensity.

The premise is that performance in the open environment will be dominated by network performance and that relatively large grained data transfers are expected to be more scalable than multiple computationally intensive interactions [19], issues that have little been studied and reported in view of emerging geographical service infrastructures. Further, it is apparent that the methods employed to define and manage quality of service in centralized mapping environments are inadequate in the open environment. Thus new methods and mechanisms are needed to specify, implement and manage quality of service in the new environments.

5 Summary

Geographical information systems and enterprizes have been under constant change in recent years with the current trend being towards Internet based distributed GI processing and dissemination. These trends raise the need for a broader concept of quality that effectively defines, captures and addresses users requirements in the distributed environments. Application of quality concepts in geographical information domains continues to be centered on the context of spatial data quality. However, quality in the contexts of business processes and geoprocessing systems are gaining growing attention.

As the technology matures and the virtual enterprize becomes the norm, future geographical information markets will increasingly be predicated on distributed infrastructures that provide guarantees on quality of services in addition to other important quality characteristics. Deliverable QoS in a distributed geographical service infrastructure is the level of performance guaranteed over that infrastructure and is characterized by a collection of values that act on performance related properties of the components that make up the infrastructure. Providing QoS guarantees in distributed Internet based infrastructures is nonetheless a big challenge which will increasingly get critical as users increase their reliance on distributed geoprocessing systems and provides motivation for evaluating the performance of emerging service infrastructures towards QoS support.

References

- [1] ISO/TC 211. *Geographic Information Services*. ISO, 2002.
- [2] Henri J.G.L. Aalders. The registration of quality in gis. In Wenzhong Shi, Michael F Goodchild, and Peter F Fisher, editors, *Proceedings of The International Symposium on Spatial Data Quality '99*, pages 23–32, 1999.
- [3] Aggrey Agumya and Gary J Hunter. Determining fitness for use of geographic information. *ITC Journal*, 2, 1997.
- [4] Nadine Alameh. *Scalable and Extensible Infrastructures for Distributing Geographic Information Services on the Internet*. PhD thesis, MIT, 2001.
- [5] Len Bass, Paul Clements, and Rick Kazman. *Software Architecture in Practice*. Addison Wesley, second edition, 2003.
- [6] Yaser Bishr. Overcoming the semantic and other barriers to gis interoperability. *Int. Journal for Geographical Information Science*, 12(4):299–314, 1998.
- [7] Yaser Bishr. Internet based large distributed geospatial databases. In *International Archives of Photogrammetry and Remote Sensing*, volume XXXIII, 2000.

- [8] Andreas Busch and Felistus Willrich. Quality management of atkis data. In *Paper presented at OEEPE/ISPRS joint workshop on Spatial Data Quality Management, Istanbul*, March 2000.
- [9] Ricardo Chalmeta and Reyes Grangel. Ardin extension for virtual enterprize integration. *The Journal of Systems and Software*, 67(3):141–152, 2003.
- [10] Guptill C.S and Morrison L.J. *Elements of Spatial Data Quality*. Elsevier Science, Oxford, 1995.
- [11] Rodolphe Devillers, Marc Gervais, Yvan Bedard, and Robert Jeansoulin. Spatial data quality: From metadata to quality indicators and contextual end-user manual. In *Paper presented at OEEPE/ISPRS joint workshop on Spatial Data Quality Management, Istanbul*, March 2002.
- [12] Coleman D.J and McLaughlin J. Defining global geospatial data infrastructure (ggdi): Components, stakeholders and interfaces. *Geomatica*, 52(2):129–144, 1998.
- [13] Mark Doucette and Chris Paresi. *Geospatial Data Infrastructure - Concepts, Cases and Good Practice*, chapter Six, pages 85–96. Oxford University Press, 2000.
- [14] Peter Fisher. Data quality and uncertainty: Ships passing in the night! In Wenzhong Shi, Michael F Goodchild, and Peter F Fisher, editors, *Proceedings of the 2nd International Symposium on Spatial Data Quality '03*, 2003.
- [15] Leonard Franken. *Quality of Service Management: A Model-Based Approach*. PhD thesis, Centre for Telematics and Information Technology, 1996.
- [16] Jorgen Giversen. Implementing the iso 19100 standards in denmarks datasets (sea-charts, top10dk and the cadastral map). In *Paper presented at OEEPE/ISPRS joint workshop on Spatial Data Quality Management, Istanbul*, March 2002.
- [17] Michael F Goodchild. Measurement - based gis. In Wenzhong Shi, Michael F Goodchild, and Peter F Fisher, editors, *Proceedings of The International Symposium on Spatial Data Quality '99*, pages 1–9, 1999.
- [18] Richard Groot. Economic issues in the evolution of national geospatial data infrastructure: A background paper for the 7th united nations regional cartographic conference for the americas, new york, usa. January 2001.
- [19] Open GIS Consortium Inc. Opengis reference model. Technical Report OGC 030-040, OGC, 2003.
- [20] ISO. *Quality Management Systems - Fundamentals and Vocabulary (ISO 9000:2000)*. ISO, 2000.
- [21] ITU-T. *Terms and Definitions related to Quality of Service and Network Performance including Dependability ITU-T Recommendation E.800*. ITU-T, 1994.
- [22] ITU/ISO. *Open Distributed Processing - Reference Model, "Part 2: Foundations" International standard 10746-2, ITU-T Recommendation X.902*. 1995.
- [23] J. Kure and F.A.A.F. Amer. Strategic challenges for national mapping agencies in developing countries. In *ISPRS*, volume XXIX Commission VI, 1992.
- [24] Ian Masser. All shapes and sizes: the first generation of national spatial data infrastructures. *Int. J. of Geographical Information Science*, 13(1):67–84, 1999.
- [25] Daniel A. Menasce and Virgilio A.F. Almeida. *Capacity Planning for Webservices: Metrics, Models, and Methods*. Prentice Hall PTR, 2002.

- [26] Douglas G. Nerbert. Developing spatial data infrastructures. *The SDI Handbook*, 2000.
- [27] Patrick Nordstrom. Large databases, large problems? multiresolution storage formats may be the answer. Geoworld, February 2003.
- [28] Gunther O. and Muller R. From gisystems to giservices: Spatial computing on the internet. In Goodchild M., Egenhofer M., and Fegeas R. Kottman C, editors, *Interoperating Geographic Information Systems*. Kluwer Academic Publishers, 1999.
- [29] Ben Chin Ooi. *Lecture Notes in Computer Science: Efficient Query Processing in Geographic Information Systems*. Number 471. Springer-Verlag, 1990.
- [30] Eurescom P806-GI. A common framework for qos/network performance in a multi-provider environment. Technical report, Eurescom, 1999.
- [31] M. Mostafa Radwan, Richard Onchaga, and Javier Morales. A structural approach to the management and optimization of geoinformation processes. Technical Report 41, OEEPE, December 2001.
- [32] Mostafa Radwan and Javier Morales. Extending geoinformation services: a virtual architecture for spatial data infrastructures. In *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume 34, 2002.
- [33] Abbas Rajabifard and Ian Williamson. Spatial data infrastructures: Concepts, sdi hierarchy and future directions. In *Proceedings of Geomatics80 Conference ,Tehran, Iran*, 2001.
- [34] Andrew B. Smith. Spatial data quality management at ordinance survey. In *Paper presented at OEEPE/ISPRS joint workshop on Spatial Data Quality Management, Istanbul*, March 2002.
- [35] Francois B Vernadat. *Enterprise Modeling and Integration*. Chapman & Hall, 1996.
- [36] Zeng Wang. *Internet QoS: Architectures and Mechanisms for Quality of Service*. Morgan Kaufmann, 2001.