

THE ARCHITECTURE OF THE GEO-INFORMATION INFRASTRUCTURE

Peter van Oosterom, Wilko Quak, Theo Tijssen and Edward Verbree

Delft University of Technology, Department of Geodesy
The Netherlands

ABSTRACT

Due to more and more exchange of geo-information within and between organisations, the need for a Geo-Information Infrastructure is also growing. In this paper the three basic components of the Geo-Information Infrastructure are described: authentic geo-data sets, geo-data processing services (geo-DBMS) and interoperability standards. The focus will be the later two, complemented with a description of the architecture of the Geo-Information Infrastructure. The OpenGIS interoperability standards are summarised and recent (geo) developments in database technology are described. Further, some advanced GIS applications based on this architecture are given: Virtual Reality (VR) and Augmented Reality (AR).

1. INTRODUCTION

Geographic Information Systems (GIS) are used within (local, regional, central) governments, utility, and other companies to support their primary business, which often depends heavily on spatially referenced data. Until recently the spatial data management was handled by the GIS software outside the DBMS. As DBMSs are being spatially enabled, more and more GISs (Arc/Info, Geomedia, Smallworld) are or will soon migrate towards an integrated architecture: all data (spatial and thematic) are stored in the DBMS. This marks an important step forward that took many years of awareness creation and subsequent system development. Many organisations are currently in the process of migrating towards this new architecture. This is a lot of work and will still take many years. The next step will be the creation of a common Geo-Information Infrastructure for related organisations; the so-called information communities. This can replace, in the long run, the exchange of copies of data sets between organisations. It requires good protocols, standardisation such as the OpenGIS [3] web mapping specification. But also the role of the spatial DBMS gets more important, because not a single organisation depends on it, but a whole community. The main use will be query oriented (and less update oriented, only the owner of the data is doing updates, others are only doing queries). An important component is the network infrastructure (bandwidth) itself. This is covered by the internet2 [4] developments and in the Netherlands by the Gigaport project [5].

The remainder of this paper is organised as follows. In Section 2 the three basic components of the Geo-Information Infrastructure are described: authentic geo-data sets, geo-data processing services (geo-DBMS) and interoperability standards. Section 3 gives an overview

of the OpenGIS interoperability standards: feature geometry models, meta data and catalog services, and Internet GIS. Recent (geo) developments in database technology are described in Section 4 with focus on the role of the geo-DBMS in the Geo-Information Infrastructure. As the query aspect is dominant in the use of external clients of geo-information and the cost of memory chips reduces, VLM (very large memory) DBMSs will become a very attractive solution. They are powerful enough to serve large number of users in the network formed by the Geo-Information Infrastructure. Further, some advanced GIS applications based on this architecture are described in Section 5: Virtual Reality (VR) and Augmented Reality (AR). Throughout the paper it becomes clear that quite a lot of research is needed to realise the Geo-information Infrastructure. Therefore, the paper concludes with Section 6, which describes our plans to set-up a research centre: the Geo Database Management Centre (GDMC).

2. THE GEO-INFORMATION INFRASTRUCTURE

The success of the Internet in general (now used intensively all over the world) has shown the power of an open infrastructure. The open (public) standards and the decentralised architecture are responsible for the many free and non-free services. At a high level the Geo-Information Infrastructure (GII) can be seen as composed of three important and quite different types of ingredients:

- Basic or authentic *geo-data sets* in different domains: topography, elevation, cadastral, geology, etc. These data sets should be well defined with respect to their data model, thematic contents, quality, accuracy, actuality, and so on.
- Geo-data processing services in general and the *geo-DBMS* specifically. The geo-data sets are maintained in these geo-DBMSs and are served to users from these geo-DBMSs via networks or traditional means. For these purposes, the DBMS has to support spatial data types and operators (simple analysis and selection oriented queries), spatial clustering and indexing (for large data sets), and if possible support for advanced analysis (topology based analysis). Also, temporal support is required in the form of some kind of future standard TSQL [15] or in the meantime through some other (non-standardised) means: extension for a specific DBMS or explicitly in the application data model.
- *Interoperability standards* are required to enable the integration of the different data sets and to combine the geo-data processing services. In fact, different organisations and individuals using each others geo-data sets in a digital environment can be regarded as parts of a distributed computing environment. One of the most obvious examples of this is an internet GIS retrieving *on-the-fly* data from different sources on the internet. In order to be able to work in such a heterogeneous world (different types of hardware, operating systems, geo-DBMSs, geo-data sets) interoperability standards at many levels are required.

All three ingredients have different aspects, which can be either technical or non-technical (organisational, financial, legal, etc.). In this paper we will focus on the recent technical developments in the area of (geo-)DBMS and interoperability standards.

One of the most time consuming tasks when implementing a GIS is obtaining geo-data. First, relevant data sets and sources have to be located, then these data sets have to be copied and converted into the local (DBMS of a) GIS. Some reasons why this process is so time consuming are that it may be difficult to find the data, the data model of the source may be very different from the model implemented by the local system, the supported exchange formats of source and destination are different. To improve this situation a lot of effort is needed in order to create a geo-information infrastructure, which has at least to cover the following aspects:

1. consensus on the geometric parts of the data model, both raster and vector data have to be supported (including different spatial reference systems);
2. how to formally describe the geo datasets (and geo processes), that is meta data standard covering both the spatial and non-spatial aspects;
3. how to access and query the meta data and how the result of such a query is returned, this is called catalog services;
4. how to select the geo-data itself;
5. how to format (and transfer) the resulting geo-data set;

Instead of always copying data sets from one system to another, a new model becomes feasible once the above aspects are covered by implementations of geo-standards. This new model allows it to keep the data at the sources, which can then be used all over the world. At the client side no data management is needed for data from other sources. This model also allows fair pricing of geo-data as every time data from the source is used (possibly through a local cache) the user can be charged for this. Currently, in the *full data-set copy* model the user has to pay for the whole data-set, even if certain data (regions) are not used at all (in a certain period). The new *data at the source* model allows fairer pricing, both viewed from the vendors and buyers point of view.

An early pilot in this area started in 1996 and was called Geoshop [18]. The following partners were involved: the municipality of Almere, a cable TV company (Casema), and the Cadastre. It was based on the C++ Magma (server) and Java Lava (client) software developed by Professional GEO Systems (PGS); see Figure 1.

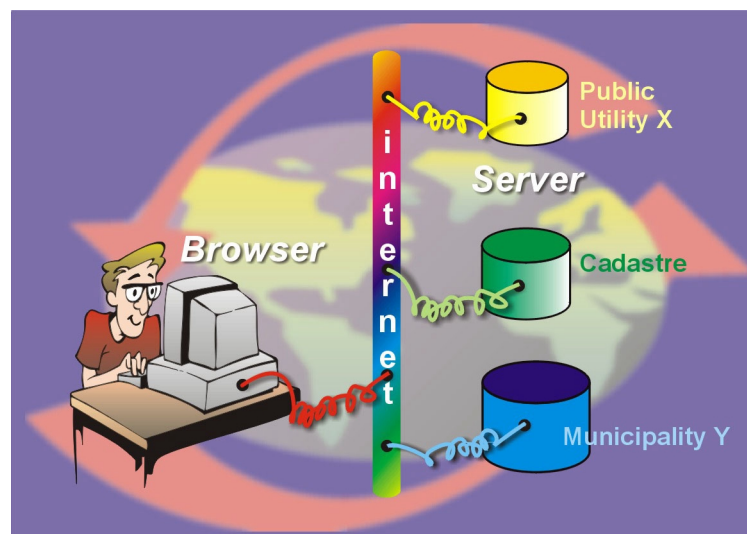


Figure 1: Geoshop example of Internet GIS.

Important characteristics are: access data from multiple and different types of sources (Ingres, Informix, DXF files), raster and vector data at server and client side, client side software platform independent (Java platform). No attempt was made to implement the payment aspects as this was expected to be a more generic issue, also treated outside the GI community. Also, no standards for querying (meta) data on returning the results were available at that time (1996).

3. OPENGIS STANDARDS

The OpenGIS consortium (OGC) and the official standardisation organisations (ISO, CEN) have addressed several aspects of the interoperating framework/puzzle. In this section we will focus on the OGC as they also (re-)use official standards when appropriate. OGC has basically two levels of standards: abstract (comparable to 'official' standards) and implementation standards. Implementation standards describe the exact interfaces (protocols) of a (part of) an abstract standard in the context of a specific distributed computing platform. An overview of the OpenGIS domain can be found in the OpenGIS Guide [16]. In Subsection 3.1 the feature geometry data model will be discussed. The next subsection covers the aspects of meta data and catalog services. Finally, Internet GIS standards are described in Subsection 3.3.

3.1 Feature Geometry

The first OpenGIS implementation standard was related to the feature geometry abstract specification. It was called Simple Feature Specification (SFS) and standardised the basic spatial types and functions. The implementation specification for the SFS are described for three different platforms: SQL[10], Corba, and OLE/COM. Currently, OpenGIS is revising/changing the feature geometry abstract specification in order to be consistent with the draft standard ISO TC 211 *Spatial schema* [8] covering also 3D types, more geometric primitives (curve and surface types), and complex features (topology). What is still missing is the implementation specification of topology for specific platforms comparable to the implementation specification for simple features. The question remains whether this would be possible with a relational DBMS using SQL (due to the *transitive closure* problem). What the abstract *feature geometry* and implementation *simple feature* specifications are for the vector model, are the abstract *earth imagery* and the implementation *grid coverage* [12] implementation specification for the raster data model.

3.2 Meta data and Catalog services

With respect to the contents and structure of the meta data, the OGC decided more or less to adopt the work by ISO TC 211 [9]. Much attention of the OGC has been paid to the related aspect of catalog services [11], describing how to access the meta data (and also data describing available computing services). This standard can and will be used in realising clearinghouses geo-information all over the world; e.g. the new NCGI in the Netherlands [17] will be based on this OpenGIS standard. An important aspect is the standard supports a decentral implementation of a clearinghouse; see Figure 2. This implies that all the meta data does not have to be stored in the central node of the NCGI, but that the queries may be passed on to decentral nodes.

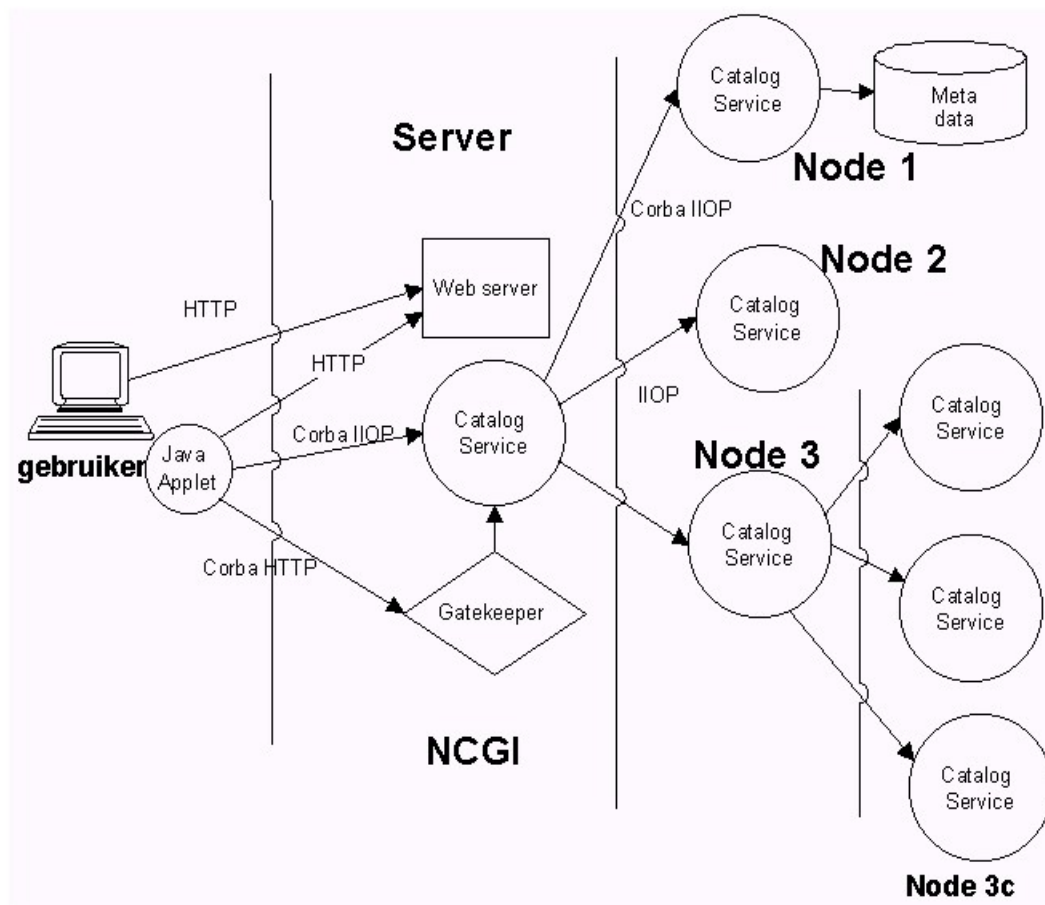


Figure 2: Architecture of the decentral NCGI (figure by Geodan IT).

3.3 Internet GIS

Strongly related to the previous standards, meta data and catalog services, are the activities in the area of Internet GIS (or web-mapping). This may be seen as an interactive (and ultimate) form of interoperability as data from multiple sources can be retrieved and combined in the web-browser. However, instead of querying and receiving meta data, now the geo-data itself is queried and received. The OGC has created in this area:

- the implementation specification *Web Map Server Interface* [14] for the query aspect using three basic functions: GetCapabilities (what is available on the server), GetMap (raster images, graphic primitives or data) and GetFeature_info (fetch attributes) and
- the recommendation *Geography Markup Language* (GML) [13] (simple features in XML with XSLT 'stylesheet' for presentation) for vector data transfer.

The query from a client gets the well-know structure of an Internet URL. The OGC has specified the name of the query parameter (e.g. BBOX, LAYERS, FORMAT) and the meaning and allowed values for these parameters. An example of the GetMap query:

```
http://b-map-co.com/servlets/mapservlet?WMTVER=0.9&REQUEST=map&
BBOX=-88.68815,30.284573,-87.48539,30.989218&
WIDTH=792&HEIGHT=464&SRS=4326&
LAYERS=AL+Highway,AL+Highway,AL+Highway&
STYLES=casing,interior,label&FORMAT=GIF&TRANSPARENT=TRUE
```

With respect to the different types of clients OGC has developed a model to compare these clients; see Figure 3:

1. thin clients (display only raster images JPEG and PNG),
2. medium clients (graphic primitives WebCGM and SVG) or
3. thick clients (data in the form of simple features XML, that is GML [5], is processed at the client side).

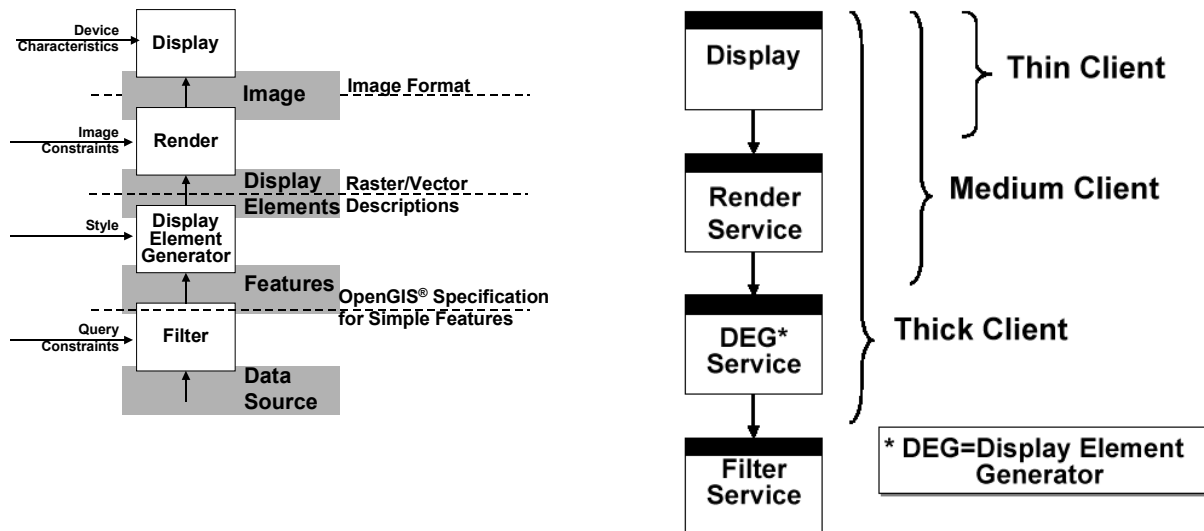


Figure 3: Different levels to exchange geo-information (taken from [14]).

It is clear that when a certain process is not done at the client side the work has to be done at the server side. For example generate graphic primitives from GML data or convert graphic primitives into images (rendering). The FORMAT parameter in the query indicates what type of result should be sent back; in the example above a GIF image is requested. This could also have been a request for GML data. Below a fragment of a GML data sets is shown, taken from [13]. The format and structure resemble the well known HTML format where everything has a begin and end: the whole object (usgs:road), thematic attributes (usgs:NumLanes) and geometric attributes (ogcgm:LineString).

```
<usgs:road ID = "1354">
<usgs:NumLanes>4</usgs:NumLanes>
<usgs:SurfaceType>LooseGravel</usgs:SurfaceType>
<usgs:Centerline ogcgm:srsName="http://www.opengis.org/srs#epsg:26751">
<ogcgm:LineString>
<ogcgm:coordinates>0.0,0.0 1.123,1.56 2.34,4.5 0.0,0.0</coordinates>
</ogcgm:LineString>
</usgs:Centerline>
</usgs:road>
```


4. DEVELOPMENT IN DATABASE TECHNOLOGY

Both developments in hardware and software in database technology will influence the future development of the Geographic Information Infrastructure. Current extensible DBMSs [1,6] are very well capable of storing spatial data. Also simple queries like zooming in and zooming out can be handled efficiently. However, more complex operations, like map overlay, on the fly generalisation, enforcing correct topology during updates etc. is still not within reach of these systems. New developments in DBMS technology that can help to build a new generation of spatial DBMSs are for example Extensible (Object Relational) DBMSs, Object Oriented DBMSs, and Very Large Memory (VLM) Databases. In the following sections we will handle these subjects.

4.1 Extensible Databases

Many DBMSs have extended their type system to include basic spatial types (such as point, line, polygon). Some databases have a built-in mechanism for extending the database, in which extension modules can be linked to the database to perform (for example) spatial operations. Between different databases the extension mechanisms differ vastly, which influence the performance and functionality of the spatial modules built with them:

Kernel Linkage vs. External Linkage

Some extension mechanisms allow execution of the extension code, directly inside the database server process. This has the advantage that the communication with the database server and the extension module is almost instant, so that extended types have almost the same performance as built-in data types. However, if there is any bug in the extension code, the complete database server may crash. This can lead to inconsistencies in the database and loss of data. Therefore, in other extension mechanisms, the extension code is run in a separate process. Communication between the database server and the extension, is done via Remote Procedure Calls (RPCs), or similar mechanisms. If the extension process crashes, the server integrity is not breached. However for some applications the cost of the RPCs might be too high and lead to an unacceptable performance loss.

Extensibility Hooks

There is a vast difference in the hooks that DBMS vendors provide for the extension modules. Sometimes it is only possible to create new datatypes. In other DBMSs, there are hooks to build new index structures, change the behaviour of the query optimizer, or even implement completely new join strategies.

4.2 Object Oriented DBMSs

As opposed to database extension, where standard databases are extended with new functionality, one can also extend a programming language with persistence. This is the approach taken by Object Oriented DBMSs. Until now, the question of whether an OODBMS is sufficiently scalable to act as an underlying database for a GIS is unanswered. Currently Jasmine is the only commercially available OODBMS supported by a mainstream IT vendor (Computer Associates).

4.3 VLM DBMSs

Over the years the memory sizes of computers has been doubled every 18 months (Moore's Law), whereas the growth in the amount of vector data has not been so fast. As a consequence, it should be possible to keep (nearly all) data inside the core memory and to perform nearly every spatial analysis operation inside the core memory of a computer. Traditional DBMSs cannot fully exploit this fact, because their architecture is geared towards solving the bottleneck between disk and memory, which is far less important in VLM Databases.

Monet is a main-memory database kernel developed at CWI, with an extension for spatial data types [2]. The system has shown excellent performance on the Sequoia benchmark [2], and currently projects are underway to extend spatial system with complex analysis functionality.

5. VIRTUAL AND AUGMENTED REALITY APPLICATIONS

The described Geo-Information Infrastructure in the previous sections, based on interoperability standards and advanced geo-DBMS technology, can be used to re-implement existing GIS applications in a more efficient manner. Further, the Geo-Information Infrastructure can be used to build new advanced applications based on the same infrastructure; for example VR (Virtual Reality) and AR (Augmented Reality). Our group has developed a DBMS VR System (Karma vI [19]). Further, we participate in the UbiCom [7] project aimed at carrying out research needed for specifying and developing wearable systems for mobile multimedia communications, that enables Augmented Reality applications such as:

- Displaying real but invisible objects, such as pipelines;
- Displaying planned or future virtual objects;
- Displaying textual information related to real world objects.

In the vision of UbiCom, a user carries a wearable terminal and a lightweight see-through display. In this display the user can see virtual information that augments reality, projected over and properly integrated with the real world. The wearable system contains a radio link that connects the user to the ubiquitous computing resources and the Internet. A camera captures the environment. Camera images are sent to the backbone and matched to a 3-D description of the real world to determine the user's position and to answer questions of the user that relate to the environment.

The vision system generates in a pre-processing phase a huge quantity (in the order of one million) of 3D-features (lines and points). The GIS technological research aims to design and build a geographic database that can store and query these and can add the features that are generated by the augmented reality system.

The specific application of this geographic database demands a high performance; when queried, a selection of the requested features must be available within the time set for an image update. Next to that the database will contain the initial environment model and the generated virtual objects (faces and textures). Another requirement is that the database will be physically separated from the augmented reality system; the wireless communication and data transfer between these two systems requires a very good compression of the geographic features. This leads to the following GIS technological research:

- Storage of 3D point, line and face features with their textures;
- Design of a 3D object database (3D data structure, 3D spatial indexing, 3D clustering);
- Querying the 3D spatial database on the basis of the augmented view while taking the initial model in account;
- Updating the database with the features generated by the vision system.

6. CONCLUSION

In the previous sections we touched on many open problems. At the Delft Technical University, the section GIS Technology will look at geo-DBMS research issues such as the standardisation of complex features (in addition to simple feature), inclusion of quality in the DBMS, 3D modelling (and visualisation tools), and advanced developments such as VLM databases. For this research the section is setting up the Geo Database Management Centre (GDMC). It will build upon the 15 years of experience of storing spatial data in the DBMS obtained within the Netherlands, the TNO research organisation, the Dutch Cadastre, and the University of Amsterdam. The University of Amsterdam contributes in the Monet project, which develops a state of the art Main Memory Database with GIS capabilities. Partners that will participate in the spatial DBMS centre of excellence at TU Delft are Sun (Java platform, Hardware), Computer Associates (Ingres, Jasmine), and Oracle (with their Oracle 8i spatial data server). Activities in this centre will include testing both performance (spatial indexing and clustering) and functionality and developing advanced spatial DBMS extensions such as support of complex spatial features (topology), 3D spatial data types and functions, spatio-temporal models, fully integrated (geodetic) quality, map generalisation functions, Internet GIS architectures (applying Java GIS applets), etc. The spatial DBMS research performed in the Centre of Excellence will help forming a fundamental basis of the GII.

ACKNOWLEDGEMENTS

We wish to express our acknowledgements to Professional GEO Systems for making their Magma/Lava software available. Further, we want to acknowledge the following Geo Database Management Centre participants: Sun Microsystems, Oracle, and Computer Associates.

REFERENCES

- [1] *ASK-OpenIngres (1994: INGRES/Object Magementent Extention User's Guide, Release 6.5. Technical report, 1994.*
- [2] *Boncz, P.A., C.W. Quak, and M.L. Kersten (1996): Monet and its Geographic Extensions: a Novel Approach to High Performance GIS Processing. In EDBT proceedings, 1996.*
- [3] *Buehler, K. and L. McKee (1998): The OpenGIS guide - Introduction to interoperable geoprocessing. Technical Report Third edition, The Open GIS Consortium, Inc., June 1998.*
- [4] *Charles, J. (1997): In the news: Internet2 promises radical change. IEEE Software, 14(6):108-109, 112, November December 1997.*
- [5] *Surfnet bulletin (September 1999): Special Issue on GigaPort.*
- [6] *Managing Spatial Data: The ESRI Spatial Database Engine for Informix (1999).*
- [7] *The UbiCom Project (1999): <http://www.ubicom.tudelft.nl/>, 1999.*
- [8] *ISO TC 211/WG 2 (1999): Geographic information - Spatial schema. Technical Report second draft of ISO 19107 (15046-7), International Organization for Standardization, November 1999.*
- [9] *ISO TC 211/WG 3 (1999): Geographic information - Meta data. Technical Report draft of ISO 19115 (15046-15), International Organization for Standardization, June 1999.*
- [10] *OpenGIS Consortium, Inc. (1998): OpenGIS simple features specification for sql. Technical Report Revision 1.0, OGC, March 1998.*
- [11] *OpenGIS Consortium, Inc (2000): OpenGIS catalog interface implementation specification. Technical Report version 1.1 (00-034), OGC, May 2000. Draft.*
- [12] *OpenGIS Consortium, Inc. (2000): OpenGIS grid coverage specification. Technical Report Revision 0.04 (00-019r), OGC, May 2000. Draft.*
- [13] *OpenGIS Consortium, Inc. (2000): OpenGIS recommendation - geography markup language (gml). Technical Report version 1.0 (00-029), OGC, April 2000.*

- [14] *OpenGIS Consortium, Inc. (2000):* OpenGIS web map server interface implementation specification. Technical Report revision 1.0.0 (00-028), OGC, April 2000.
- [15] *Snodgrass, R.T., I. Ahn and G. Ariav (1994):* Tsql2 language specification. SIGMOD Record, 23(1):65-86, 1994.
- [16] *Tuijnman, F. and P.J.M. van Oosterom (1998):* Geoshop, een multi-server internet-GIS (in Dutch). Geodesia, 40(4):165-174, April 1998.
- [17] *Absil van de Kieft, I. and B. Kok (1997).* The development of a geo meta data service for the Netherlands. In Third Joint European Conference & Exhibition on Geographical Information (JEC-GI'97), pages 1165-1176, April 1997.
- [18] *Berg C. van den, F. Tuijnman, T. Vijlbrief, C. Meijer, H. Uitermark, and P. van Oosterom (1997):* Multi-server internet GIS: Standardization and practical experiences. In International Conference and Workshop on Interoperating Geographic Information Systems, Santa Barbara, California, USA, December 3-4 and 5-6, 1997, interop'97, pages 365-378, Boston, 1997. Kluwer Academic Publishers.
- [19] *Verbree, E., G. van Maren, R. Germs, F. Jansen, and M.-J. Kraak (1999):* Interaction in virtual world views - linking 3D GIS with VR. International Journal of Geographic Information Science, 13, June 1999.

CO-ORDINATES

Prof.dr.ir. P.J.M. van Oosterom

Drs. C.W. Quak

Drs. T.P.M. Tijssen

Ir. E. Verbree

Delft University of Technology

Faculty of Civil Engineering and Geosciences

Department of Geodesy

Section GIS technology

Thijssseweg 11

2629 JA DELFT

The Netherlands

Tel. +31-15-278 6950

Fax +31-15-278 2745

E-mail oosterom@geo.tudelft.nl

c.q.quak@fgeo.tudelft.nl

t.p.m.tijssen@geo.tudelft.nl

e.verbree@geo.tudelft.nl

URL [//www.geo.tudelft.nl/GIS/gist_e/index.htm](http://www.geo.tudelft.nl/GIS/gist_e/index.htm)