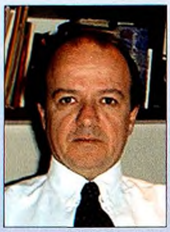
Article



A Model for the Generalisation and Transfer of Linear Features over the Web

By Dr Eng. Andriani Skopeliti, Cartographer, and Assoc. Professor Lysandros Tsoulos, NTUA Cartography Laboratory, Faculty of Rural and Surveying Engineering, National Technical University of Athens, Greece

Abstract

The evolution of open standards and especially those pertaining to the family of XML technologies have a considerable impact on the way the spatial data are stored, transferred, processed and displayed. This paper elaborates on the development of a service for the transfer and dynamic generalisation of linear features over the Web. This entails the parametric description of lines' shape with a number of measures, the segmentation of lines in homogeneous segments and the development of the relevant application schema. The methodology can be utilised for the generalisation and transfer of linear features in the framework of the IHO Transfer Standard [S-57] and its representation in GML.

Résumé

L'évolution des normes ouvertes et notamment de celles qui appartiennent à la famille des technologies XML ont un impact considérable sur la manière dont les données spatiales sont stockées, transférées, traitées et affichées. Cet article se réfère au développement d'un service pour le transfert et la généralisation dynamique des éléments linéaires sur le Web. Ceci inclut la description paramétrique des formes de lignes avec un certain nombre de mesures, la segmentation des lignes dans des segments homogènes et le développement du schéma d'application approprié. Cette méthodologie peut être utilisée pour la généralisation et le transfert des éléments linéaires dans le cadre de la norme de transfert de l'OHI (S-57) et de sa représentation en GML.

Resumen

La evolución de normas abiertas y especialmente aquellos pertenecientes a la familia de la tecnología XML tienen un considerable impacto en la forma que los datos espaciales son guardados, transferidos, procesados y desplegados. Este artículo se refiere al desarrollo de un servicio de transferencia y generación dinámica de objetos lineales en la Web. Esto incluye la descripción para-métrica de las formas de líneas con un número de medidas, la segmentación de líneas en segmentos homogéneos y el desarrollo del relevante esquema de aplicación. La metodología puede ser utilizada para la generalización y transferencia de objetos lineales en el marco de la Norma de Transferencia de la OHI (S-57) y su representación en GML.

Introduction

Recent developments in Web mapping and on-demand mapping have led to the development of prototype systems for the transfer and distribution of digital data. To generate such maps/charts the corresponding data sets must be available at a specific scale, something rare for the majority of requests. There are two alternative approaches to the problem at hand: the first is based on the existence of a multi scale spatial database. This approach has a major disadvantage; the arduous task to keep all data up to date independently (Hardy, 2000). In the second approach, data sets at different scales are derived from a database utilising a generalisation module incorporated in a web map service. The Open Geospatial Consortium (OGC) in the OGC Web Services (OWS) initiative provides for the 'feature generalisation service'. More specifically, the OGC includes the 'feature generalisation service' in the geographic processing services characterised as spatial. It is described as a 'service that reduces spatial variation in a feature collection to increase the effectiveness of communication by counteracting the undesirable effects of data reduction' (OGC and ISO, 2001).

A Web Map Service, as an on-demand mapping suite, includes a generalisation module that handles on the fly generalisation. On-demand mapping is demand-driven, that is, map products are generated according to user needs. On the fly generalisation is the real time creation, upon user's request, of a cartographic product appropriate for a specific scale and purpose. On the fly generalisation is a time-critical task and algorithms suitable for this purpose must be fast and effective and supported by pre-computed data structures and attributes (Cecconi, 2003). This is accomplished by metadata that can be provided at three levels within an exchange set (IHO S-57, 2000) to make it more useful for a properly defined application. For on-demand mapping the metadata contribute to the resolution of a number of problems inherent to generalisation. The better the metadata content is tuned to the requirements of the defined application, the more the required transfer and generalisation processes are enhanced. Meta-information about the semantics and the geometry of objects or groups of objects (patterns, structures) is critical for successful generalisation (Jones et al., 1996).

The description of the nature and the shape of linear features has been a challenging research issue in contemporary cartography. Recent research resulted in a number of parameters (measures), which describe qualitatively/quantitatively the line shape. A new approach is introduced here, which tests and enhances the existing results utilising a large set of geographical data comprised of a wide variety of line types and a comprehensive list of the measures identified in the literature. The set of parameters adopted are selected from a broad set of measures utilising Principal Components Analysis. Based on the parametric description of line shape, segments are subsequently allocated into similar shape groups through Cluster Analysis.

The generalisation of coastline, contours and other linear features displayed on a nautical chart is a critical task in the production of the paper or electronic form of the chart. Most operators used in generalisation of linear data such as simplification, smoothing etc. influence positional accuracy that is further analysed to horizontal position accuracy and shape fidelity (European Committee for Standardization, 1996). When generalised data are exchanged, all parties involved must be aware of the data quality. It is obvious that the transfer of digital data should be supported by metadata describing - amongst others - positional accuracy.

It is quite common that data downloaded from a database or contained in an exchange set, must be generalised in order to be used for spatial analysis or map/chart composition projects at a scale smaller than the source scale. The cartographer's choices in generalisation are strongly related with line shape. As many researchers have pointed out, cartographic generalisation would be successfully implemented, if it were based on the knowledge of the line shape. A generalisation service will perform much better with meta-information describing the shape of linear features. Measures can be used to detect and evaluate particular characteristics of a linear feature in order to select the optimum generalisation approach/algorithm and enable the qualitative assessment of the results (Skopeliti, 2001).

This paper addresses the design and implementation requirements of a service for the transfer and generalisation of linear features over the Web. The next section elaborates on a method for shape description and structure recognition of linear features and introduces measures that describe shape change of a feature with respect to its original form. It is followed by the development of a model for linear data generalisation that is implemented through a GML application schema that encodes the generalisation information and a prototype system architecture supporting the Web service. The paper concludes with a review of the contribution of open standards and the developed Web service to the transfer and representation of geospatial information.

Line Shape Description and Structure Recognition

In the course of their research on the subject, the authors developed a methodology for the parametric description of the shape of linear features (Skopeliti and Tsoulos, 1999; Skopeliti, 2001). This work has been re-examined (Tzamakou, 2004), utilising the complete set of measures/parameters found in the literature and an extended data set, in order to make the results applicable to a broader application area such as data transfer and generalisation over the Web. The fundamental concept in this approach is the use of measures, which are calculated at different levels of constant resolution. This concept has been adopted by a considerable number of researchers, who followed different approaches for line shape description (Buttenfield, 1991; Bernhardt, 1992; Mokhtarian and Mackworth, 1992; Mandelbrot, 1967). In particular for angular measures computations, like in Carstensen (1990), Thapa (1989) and Plazanet (1996), lines are preprocessed in order to acquire a common resolution taking into account the source scale. This way, angularity calculations are comparable with no bias due to the vertices spacing.

The measures evaluated are those proposed by Buttenfield (1991), Bernhardt (1992), McMaster (1986) and Jasinski (1991) and they are listed in Appendix I. The study of those measures' ability to describe the line shape requires a data set with line segments varying widely in shape. In the framework of the project natural linear features (coastline, streams) were used since their shape cannot be described by basic geometric archetypes. In order to ensure the required variety in shape, data were derived from an extended geographic area and selected from a cartographic database at a medium scale. At a scale such as 1:50,000, data have been generalised but still preserve their main natural/structural characteristics. In addition, at this scale the lines' shape (i.e. smooth or sinuous) can be recognised by the map reader. Moreover, this scale is used by the majority of national mapping agencies as the reference source scale of a national dataset where from other cartographic products are derived.

In this study data derived from National Oceanic and Atmospheric Administration (NOAA) Medium Resolution Digital Vector Shoreline (NOAA, 2003) are used. NOAA's Medium Resolution Digital Vector Shoreline is a high-quality, Geographic Information System-ready, general-use digital vector data set created by the Strategic Environmental Assessments (SEA) Division of NOAA's Office of Ocean Resources Conservation and Assessment. Compiled from hundreds of NOAA's coastal charts this product comprises over 75,000 nautical miles of coastline at 1:40,000 scale (nearly 2.5 million vertices), representing the coterminous US. The selection of the data segments that make the data set is guided by their shape and length. Segments are selected in order to bear the required variety in shape and are characterised as very smooth, smooth, sinuous and very sinuous. A minimum line length is required for the calculation of the fractal dimension parameter that also allows for an identifiable line shape e.g. a minimum number of bends in order to create a sinuous form. Finally the size of the data set should be sufficient to support statistical processing. To this effect, 133 line segments have been selected with length between 5,000 and 7,000 metres.

Statistical Analysis

In general, factor analysis is utilised to examine the underlying patterns or relationships of a number of variables and to determine whether the information can be condensed or summarised in a smaller set of factors (Hair et al., 1998). In statistics factor analysis is carried out through Principal Components Analysis (PCA). Principal Components Analysis is performed on the above-mentioned measures, which are computed for all the linear segments in the data set. Based on the 'scree' test and the 'latent root' criterion three factors are retained, which represent 91% of the variance of the tested measures. Taking into account the PCA results and cartographic judgment, a subset of measures for line complexity description is formed. The 'average magnitude angularity', the 'error variance' and the 'average angularity' are used for the parametric description of line shape and the clustering of line segments into groups of similar complexity. A detailed description of these measures can be found in Appendix II. The measures selected exhibit low correlation between them and - therefore - they can be used in a cluster analysis procedure. Cluster analysis is a statistical process used for the classification of the line segments based on the measures chosen through factor analysis. For hierarchical cluster analysis, the Ward method and the Euclidean distance (Hair et al. 1998) produce the best classification results. Based on cluster analysis the dataset is successfully classified into four groups that can be characterised as smooth, smooth with a strong arc, sinuous and very sinuous (Figure 1).

In order to achieve a successful classification for linear entities, lines should be homogeneous throughout their extent. The development of a segmentation methodology is therefore a prerequisite for successful classification. Such a methodology should identify 'homogeneous' parts and differentiate parts with varying degree of complexity. A considerable number of researchers have addressed the problem of linear features' segmentation (Plazanet et al., 1995; Wang and Muller, 1998; Dutton, 1999). The authors developed a methodology for natural linear features partitioning into homogeneous segments based on line shape assessment, utilising fractal dimension and the above mentioned methodology for the description of lines' shape (Skopeliti and Tsoulos, 1999; Skopeliti, 2001).

Generalisation Strategy and Assessment of Generalisation Quality

The assessment of generalisation quality of linear features, calls for the quantitative description of the resulting horizontal position and shape. Shape change of a linear feature with respect to its original form, can be assessed through the description

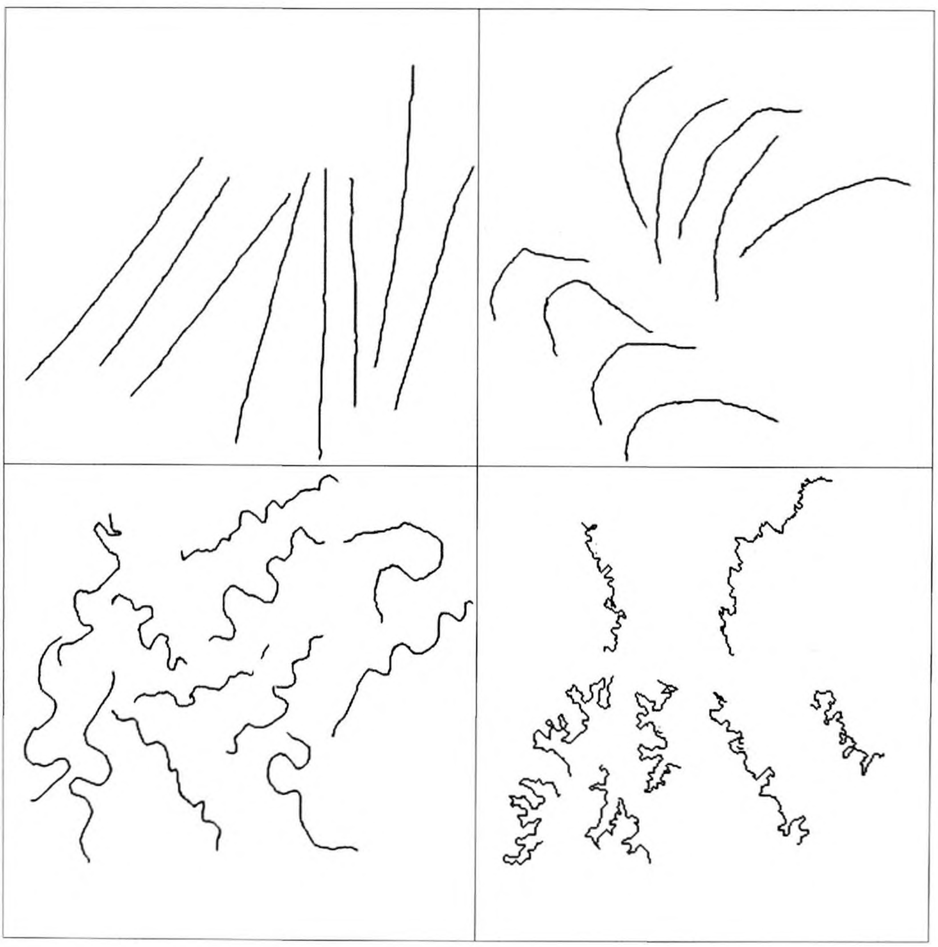


Figure 1: Classification of a subset of the data used in 4 groups.

of resulting line shape with the same group of measures. In cluster analysis, the distance between two lines in the parameters' space implies similarity. The distance between the original and the generalised line implies shape change. This is a quantitative assessment of shape change due to generalisation. A number of measures for the assessment of positional deviation between the original and the generalised line have been identified by the research community:

- the average Euclidean distance from the original to the generalised line or from the generalised to the original line, the Hausdorff distance (Abbas et al., 1995)
- the ratio of the area between the original and the generalised line to the length of the original line (McMaster, 1987) etc.

In this study the Hausdorff distance, which describes the maximum error, is used. In previous studies (Skopeliti, 2001; Skopeliti and Tsoulos, 2000; Skopeliti and Tsoulos, 2001a; Skopeliti and Tsoulos, 2001b), the authors have showed that besides the assessment of the generalisation quality of linear features, the same measures can be used for the 'automatic' identification of the most appropriate generalisation strategy with respect to quality.

Experiments with alternative generalisation strategies using different algorithms, varying sequence of generalisation operators and values of parameters, resulted in the selection of the most appropriate algorithm for each category of linear features. A generalisation strategy is considered optimal when a number of quality criteria and constraints posed by the cartographer (preservation of shape, global complexity, original lines' characteristics, degree of simplification suitable for the new map scale, minimisation of shape distortion, minimisation of horizontal position error etc.) are satisfied. The knowledge acquired from the most appropriate generalisation strategy with respect to positional error minimisation can be stored as metadata in the database.

The aforementioned methodology constitutes the framework for the development of an enhanced model for linear data generalisation. This model is comprised of the following elements:

- Structure Recognition: linear features are segmented in homogeneous lines
- Shape Description: linear features are classified into similar shape groups characterised as 'sinuous', 'smooth' etc.
- Procedural Knowledge on Generalisation: the optimal generalisation strategy is recommended for groups of lines with similar shape
- Generalisation Quality Assessment: use of measures that describe line shape change and horizontal position deviation between the original and the generalised lines.

Data and Metadata Transfer Schema

The ultimate goal of this work is the development of a model for the encoding of geometry and attributes of linear features that allows for efficient transfer and on the fly generalisation. This enhanced model is implemented with the use of XML family of technologies. GML is an XML data encoding scheme for the transport and storage of geographic information including both the spatial and the non-spatial properties of geographic features. GML has been designed to uphold the principle of separating content from presentation. It provides mechanisms for the encoding of geographic feature data regardless of the way the data are presented to a human reader.

GML has revolutionised the treatment of spatial information and is becoming a de-facto standard for the transfer and representation of geospatial information. Several reasons led to the widespread use of the GML standard (GML, 2004):

- It is based on a common model for geography that is the OGC Abstract Specifications
- It is based on XML. As a result it is a text based encoding scheme, human readable and platform independent, suitable for web applications and transfer through the Internet. In addition, the XML technology facilitates the integration of GML data with non-spatial data e.g. with statistical data, which are also encoded in an XML-based standard
- It has been adopted by the vast majority of GIS vendors

GML successfully fulfills the requirements of a spatial data transfer protocol as stated by the International Cartographic Association. It also meets the three levels of connectivity that must be satisfied at physical, logical and semantic level for geographic data transfer. The OGC, an industry funded consortium, is interested in the development and support required for the transition of this specification from a *de facto* to a *de jure* status. The key concepts used by GML to model the world are drawn from the OpenGIS Abstract Specification and the ISO 19100 series. GML 3.1 provides a variety of objects for the description of geography including features, coordinate reference systems, geometry, topology, time, units of measure and generalised values. The current version of GML addresses the following topics that were not or inadequately addressed in GML ver. 2, which are reflected in the corresponding base schemas (GML, 2004):

- Represents geospatial phenomena in addition to simple 2D linear features, including features with complex, non-linear, 3D geometry, features with 2D topology, features with temporal properties, dynamic features, coverages and observations
- Provides more explicit support for properties of features and other objects with complex values
- Represents spatial and temporal reference systems, units of measure and standards information
- Uses reference system, units and standards information in the representation of geospatial phenomena, observations and values
- Represents default styles for feature and coverage visualisation
- Conforms with the ISO 19100 series of standards.

An application schema declares the actual feature and property types of interest for a particular domain using components from GML in standard ways. Broadly speaking, these involve the definition of application-specific types, which are derived from types provided in the standard GML schemas or direct inclusion of elements and types from the standard GML schemas. The base GML schemas effectively provide a meta-schema or a set of foundation classes from where an application schema can be implemented.

The enhanced model is implemented in GML in order to be used for linear data generalisation and vector data transfer over the Web. As every GML application schema, it uses predefined GML classes. Figure 2 shows the UML diagram that describes the proposed model.

The first class in this model is the root element, which is the element embracing all the other elements in the GML document. This class is called *SpatialData* and is an Abstract Feature Collection. The second class, called *ThemeOfLinearData*, is an Abstract Feature that corresponds to any geographic theme that uses linear data such as coastline, streams etc. This class has a number of attributes: *theme_description, theme_scale and theme_date* that are used to store a short description of the geographic data, the source scale and the creation date. Another class is the *GroupOfSimilarLines*, which is an AbstractFeature. This class is used to represent a group of linear segments with similar shape. From this class a number of instances representing a group of lines with specific shape can be generated e.g. sinuous, very sinuous etc. as one can see in the UML object diagram (Figure 3). The *GroupOfSimilarLines* class has a complex attribute called *GeneralisationStrategy* that carries

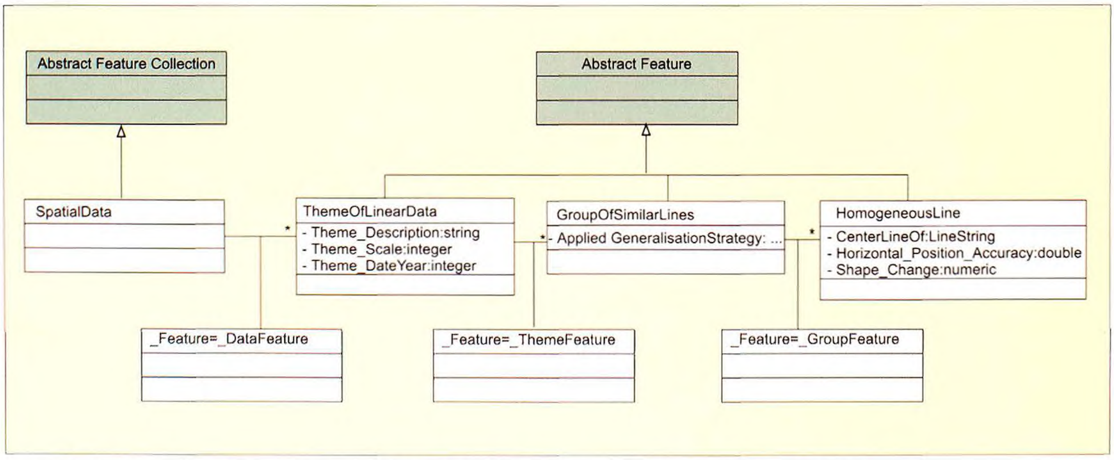


Figure 2: UML Class diagram of the enhanced model for linear data generalisation encoding.

information concerning the generalisation of this group of lines. At present this information stores the generalisation operator used, the algorithm applied and the values of parameters needed by the algorithm. The last class of the model is the *HomogeneousLine*, which is an abstract feature used to store the geometry of lines utilising the gml:centerLineOf:LineString element. Additionally, two attributes are used to describe the positional accuracy: *HorizontalPositionAccuracy* that is used to report the value of the measure used for the assessment of the positional deviation between the original and the generalised line and the *ShapeChange* attribute that is used to report the value of the measure used for the assessment of the shape change between the original and the generalised line. For better understanding of the UML Class diagram of the enhanched model for linear data generalisation (Figure 2), a UML Object diagram showing instances of the classes for linear data generalisation is provided (Figure 3).

The IHO publication S-57 describes the standard to be used for the exchange of digital hydrographic data between national hydrographic offices and for its distribution to manufacturers, mariners and other data users. The United Kingdom Hydrographic Office (UKHO) in association with Galdos Systems Inc has produced an initial version of a GML schema for ENC data. This schema is intended to aid in the representation of S-57 Edition 3.1 Electronic Navigational Charts (ENCs) in GML. The schema will aid the interoperability of data produced by different vendors and facilitate the representation of S-57 ENC data in GML. S-57 Schemas are publicly available at: http://www.ukho.gov.uk. The GML application schemas for S-57 ENC consist of the following XSD files (Burggraf, 2004):

- Objects.xsd: covers all valid S-57 ENC objects except C_ASOC and the cartographic objects
- Attributes.xsd: covers all valid S-57 ENC attributes except cartographic attributes
- AbstractAndSuperTypes.xsd: contains all abstract base types, top-level feature collection and corresponding metadata
- SupportTypes.xsd: contains supporting simple types that are shared by the S-57 schema elements.

The Objects.xsd represents the S-57 Object Catalogue (roughly 200 Objects) including the Coastline feature as a *CoastlineType* element. This *CoastlineType* element can be enriched in order to provide for each coastline homogeneous segment the

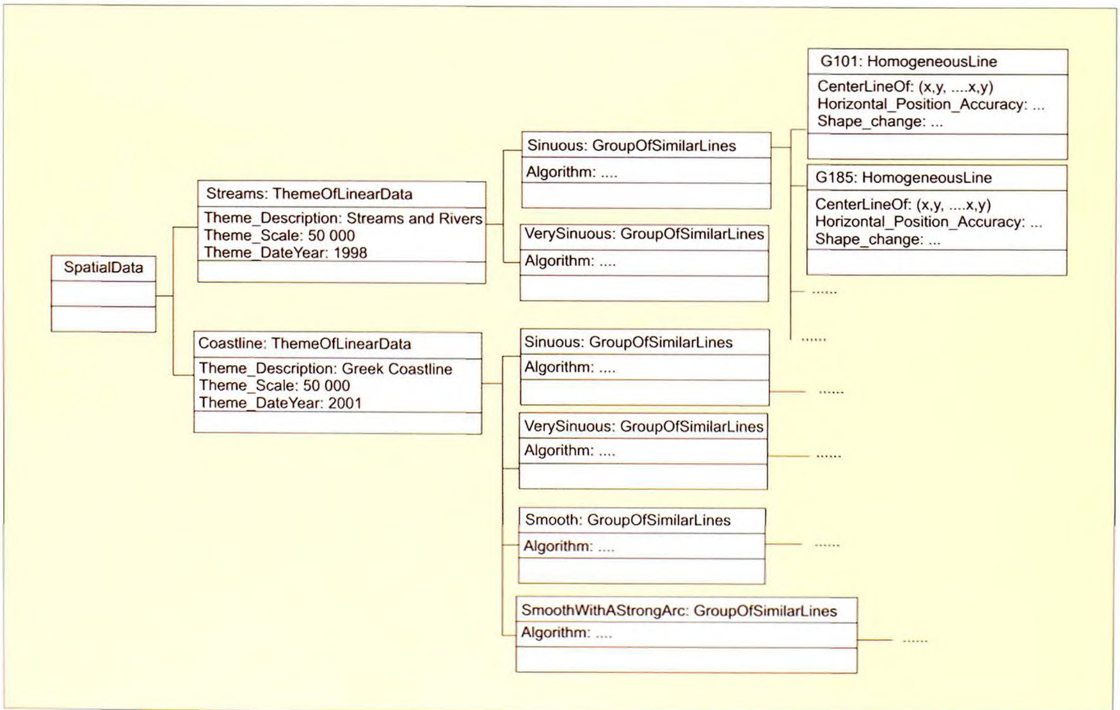


Figure 3: UML Object diagram showing instances of the classes.

information supported by the enhanced model for linear data generalisation (Figure 4). This is achieved through the Generalisation_info attribute, which is a *GeneralisationInfoType* element. The *GeneralisationInfoType* element has five attributes: *Shape_Description* that is used to classify and characterise the linear segment according to its shape, *HorizontalPositionAccuracy* that is used to report the value of the measure used for the assessment of the positional deviation between the original and the generalised line, the ShapeChange attribute that is used to report the value of the measure used for the assessment of the shape change between the original and the generalised line and the *Generalisation_algorithm* and *Generalisation_operator* that are used to report the optimum algorithm and operator to be used.

Data Transfer and Generalisation Service Prototype

In this section, the architecture of a prototype system for on the fly generalisation service is presented. The system is modular and complies with the specifications proposed by the Open Geospatial Consortium (OGC) such as: Web Map Service (WMS, 2004), Geographic Markup Language (GML, 2004), Web Feature Service (WFS, 2004) and standards of the XML family of technologies recommended by the World Wide Web Consortium (W3C) such as: eXtensible Stylesheet Language Transformation (XSLT 1.0, 1999) and Scaleable Vector Graphics (SVG 1.2, 2004). These technologies are used successfully in other systems designed for real-time mapping such as GiMoDig (GiMoDig, 2004) and GEMURE (GEMURE, 2004).

GML and SVG standards are complementary: GML is used for the storage and distribution of geographic data and SVG is used for data presentation. SVG is a two-dimensional vector graphics format written in XML. As an XML format, SVG is platform independent and non-proprietary. The most important advantages of SVG are (Watt, 2002):

- SVG is searchable. Text and elements within an SVG file may be indexed by the search engine software. Using a search engine, users may look for a map that has a particular name on it and they can search that map for its location. The equivalent search with a raster format requires optical character recognition software
- SVG graphics are resolution and device independent
- SVG format creates smaller size files that reduce download times compared to bitmapped graphics. That is why they are better suited for devices with low bandwidth and limited memory
- SVG images can be panned and zoomed without degrading image quality

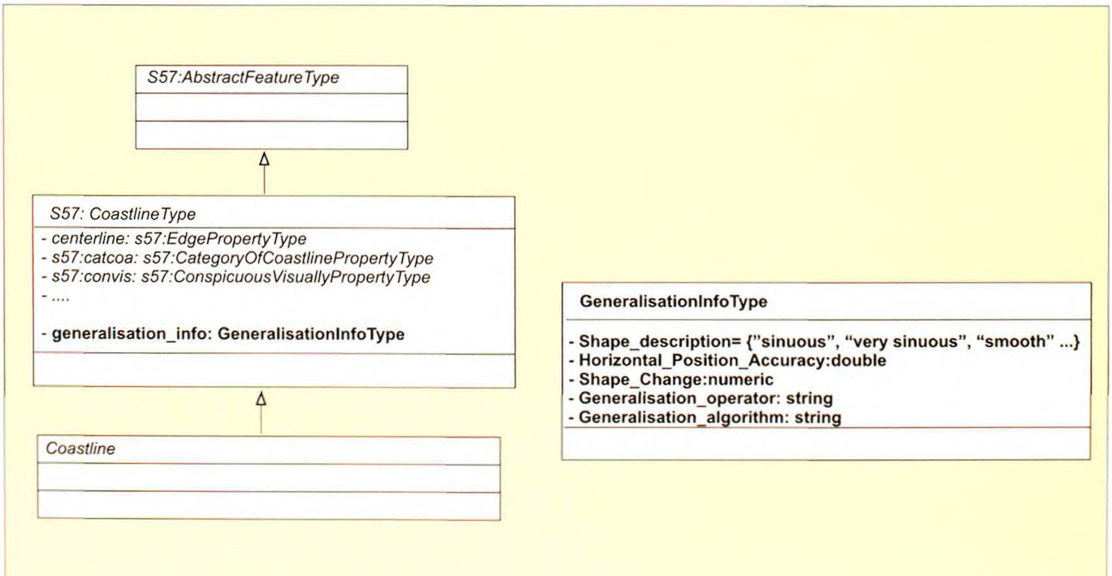


Figure 4: UML Class diagram of the S-57 GML schema enriched with generalisation metadata.

- SVG specification is fully compatible with existing technologies like HTML and XHTML, XLink, XML Namespaces, DOM, CSS and XSL
- SVG graphics can use scripting to provide interactivity and animation

In most applications geodata are stored in a database and when a map request is executed a GML-encoded file is generated. The GML file is subsequently translated into SVG format through the *eXtensible Stylesheet Language Transformation* (XSLT). XSLT provides a set of tools for transforming documents from their original structure to a new structure and is commonly described as a transformative style language (DeCharme, 2002). Instead of adding information to the original document structure, it creates a new document structure (*stylesheet*) based on rules applied to the content of the original.

An XSLT *stylesheet* is a collection of template rules. Each template rule includes a pattern that identifies the source tree nodes to which the pattern applies and a template that is added to the resulting tree when the XSLT processor applies a specific template rule to a matched node. For example, the same XML document can be transformed to GML, SVG or HTML by applying to the source document three different stylesheets. It can be considered as a programming language for the process and transformation of XML documents. It is affiliated with *XML Path Language* (XPath), which is a specialised language for addressing parts of an XML document.

The OGC Web Map Service (WMS) is based on the OGC WMS specifications and the ISO/TC211 specification (ISO 19128). It provides three operation protocols in support of the creation and display of registered and superimposed map-like views of information that come simultaneously from multiple sources, which may be remote and heterogeneous, in standard image formats such as Scalable Vector Graphics (SVG), Portable Network Graphics (PNG), Graphics Interchange Format (GIF) or Joint Photographics Expert Group (JPEG). Through WMS, users can submit requests in the form of Uniform Resource Locators (URLs) by using a standard web browser (OGC document 04-024 2004). The Web Feature Service (WFS, 2004) is an OGC specification for describing data manipulation operations at the feature level on OpenGIS simple features e.g., points, lines, and polygons (OGC document 02-058 2002). WFS is written in XML and uses GML to represent features. The database - or datastore in OGC's terms - used to store geographic features can be in any format but it should be opaque to client applications. Users can send a request in XML to a WFS server; WFS, which connects with various formats of databases, processes the requests and sends the response In XML back to users. WFS uses a subset of XPath expressions for referencing properties and HTTP as the distributed computing platform. To support transaction and query processing at feature level on datastores, five operations are defined in OGC WFS: GetCapabilities; DescribeFeatureType; GetFeature; Transaction and LockFeature. It is this feature-level data manipulation that allows users to download only the desired data instead of the whole dataset and makes WFS valuable for time-critical applications because it can reduce considerably the data access time and integration and thus significantly improve the speed of decision-making (Peng and Zhang, 2004).

One of the main elements of the prototype system for a data transfer and generalisation service over the Web is the geographic database. The enhanced linear geographic data model is implemented in the geographic database. Linear features are segmented in homogeneous segments, the segments are classified into similar shape groups and knowledge of the most appropriate generalisation strategy in relation to line shape is provided. The spatial database is implemented in Oracle 9i, an object-relational Data Base Management System (DBMS). In the current version of Oracle 9i the object-relational model has been extended to include an SDO_GEOMETRY type used to store the geometry of the objects. The geometry is stored using what the OGC calls 'simple features' and 'composite collections' of these primitive types. The same standards are also used in GML and in the database conceptual model. This compatibility is enviable, since the geo-database is developed to work as a repository from where the GML data are generated. Another advantage is Oracle's ability to communicate with well-known GIS systems such as ESRI's Arc/Info. This factor facilitates the population of the database and spatial data editing.

Loading data into the spatial database is a straightforward process with the use of ArcSDE,

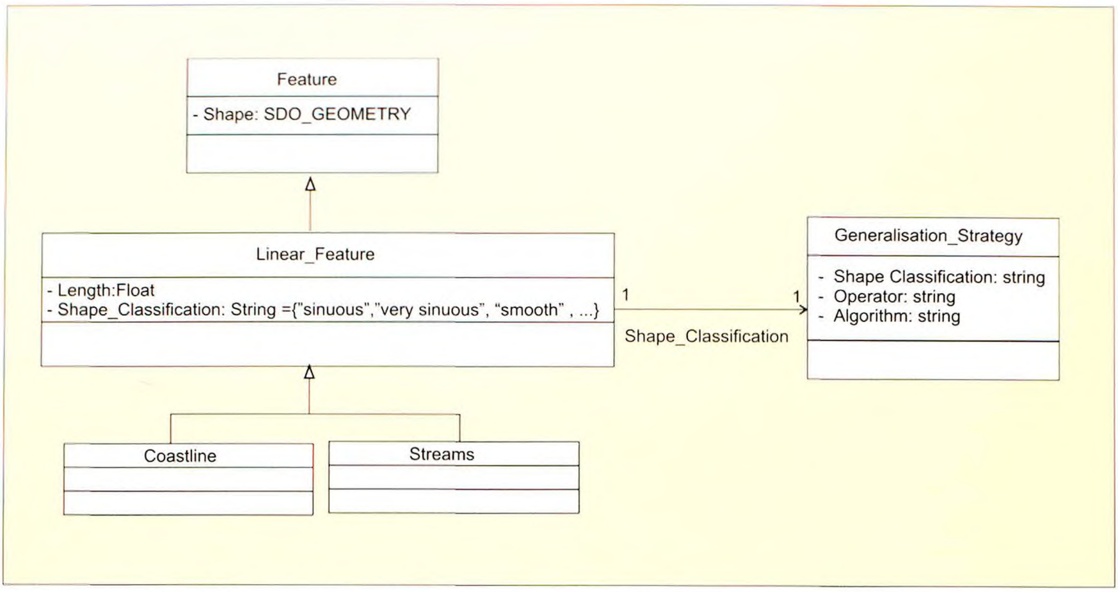


Figure 5: UML class diagram of the Geographic Database: linear features.

which is an Arc/Info client that works as a gateway between Arc/Info and Oracle. For the creation of the spatial database each layer stored as an Arc/Info coverage is converted to a Feature Type. The Feature Type preserves the geometry and all properties appearing in the feature attribute table of the coverage. A mapping exists between the geometry type of features in the geo-database and the coverages. In addition a new field called Object_ID is created for all geometry types, which guarantees a unique ID for each object as required by the data model. All Info tables that store additional properties of the features are transformed into Oracle tables. Coverage and Info table items are mapped based on a combination of their type and their width. The elements of each linear fea-

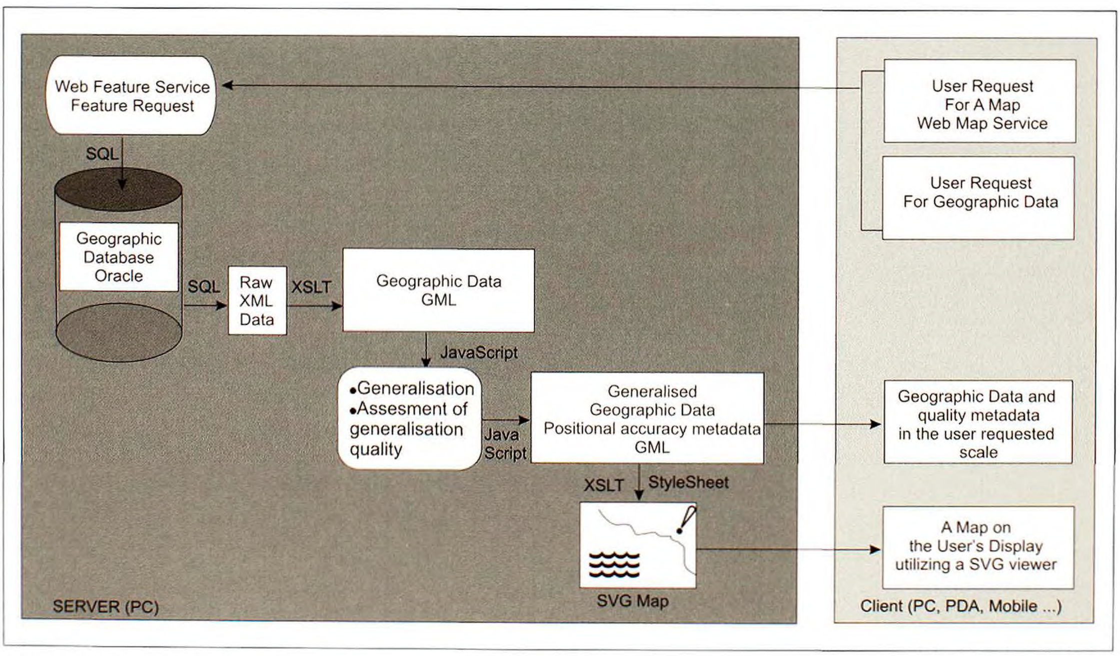


Figure 6: Prototype System Architecture.

ture are stored in the database segmented in homogeneous line segments.
The 'Shape_Classification' attribute stores the information that classifies each one of them in a group of lines with similar shape such as 'sinuous', 'smooth' etc. An additional table 'Generalisation Strategy' stores (in field 'Algorithm') for each group of lines the optimal algorithm e.g. Douglas – Peucker, Lang etc. and for each generalisation operator e.g. simplification, smoothing etc. (in field 'Operator') the proper operator. At the moment and based on the knowledge on the optimal generalisation solution in relation to line shape and quality constraints acquired through the authors' experiments (Skopeliti, 2001; Skopeliti and Tsoulos, 2000; Skopeliti and Tsoulos, 2001a; Skopeliti and Tsoulos, 2001b) information on algorithms is available for line simplification. The UML class diagram of the Geographic Database in Figure 5 shows the storage of the linear features that supports the enhanced model for linear data generalisation.

The system uses the Open GIS Specification for Web Map Services (WMS, 2004) for the communication between the server and the client. A map request is formed by the user in the client, based on WMS and is subsequently transferred to the system server via Simple Object Access Protocol (SOAP, 2003). This request is transformed in a WFS request to the database (Figure 6). Data for the linear feature requested are extracted through a spatial query from the database in XML raw format. The elements of each linear feature are stored in the database segmented in homogeneous line strings along with the information that classifies each one of them in a group of lines with similar shape. The XML data are converted to GML format utilising XSLT (Tsoulos et al., 2003) in accordance with the aforementioned GML application schema. A JavaScript handles the generalisation of linear features applying the algorithm suitable for this category of lines based on the metadata in the GML file and a new GML file that stores the generalised data is created. In addition, the Java module calculates the values of the measures used to describe the line shape change and the horizontal position deviation. This information appears as metadata along with the data in the new GML data files. The graphical output of the generalised data is a map in SVG generated through an XSLT transformation (Spanaki, 2002; Antoniou, 2004; Spanaki et al., 2004).

Conclusion

This paper addresses the development of data transfer and generalisation service over the Web utilising state-of-the-art standards and technologies introduced by the Open Geospatial Consortium such as: Geographic Markup Language for the encoding and transfer of linear data and metadata, Scalable Vector Graphics for map display, Web Feature Service for data retrieval from the database and Web Map Service for map query and transfer. This service applies on the fly generalisation to the linear data retrieved from a database based on the enhanced geographic data model. This model is based on the following assumptions: linear features are segmented in homogeneous segments, segments are classified into similar shape groups and 'a priori' knowledge of the most appropriate generalisation strategy in relation to line shape is provided. On the fly generalisation uses measures, which describe line shape and the horizontal position deviation between the original and the generalised line, to assess generalisation quality. Generalised data and metadata are encoded according to the GML schema proposed. The user through the client may access the generalised data and the quality metadata or a map in SVG. The proposed service is an integral approach to dynamic transfer and generalisation of linear features based on open standards that constitute a modern framework in geodata transfer and processing.

Appendix I – List of Measures Evaluated

Fractal dimension; bandwidth; segmentation; error variance; concurrence; average angularity; standard deviation of average angularity; average magnitude angularity; standard deviation of average magnitude angularity; average positive angularity; average negative angularity; average angularity length; average magnitude angularity length; average curvilinearity length; standard deviation of curvilinearity length; curvilinearity ratio; average vector displacement from baseline; average magnitude vector displacement from baseline; ratio of the average vector displacement from baseline to the average magnitude vector displacement from baseline and Jasinski' s angularity. The ratio of the line length to the base line length is also introduced. This measure describes the deviation of the line from its simplest form, that is, the base line.

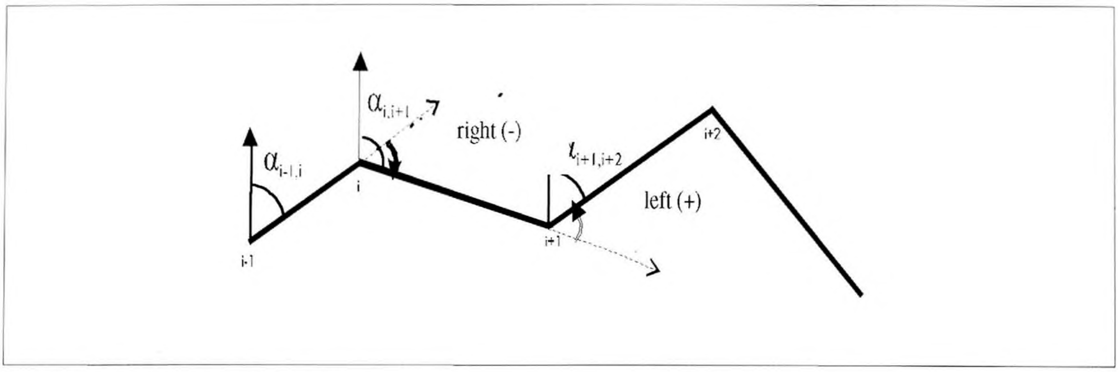


Figure 7: Angularity sign definition.

Appendix II - Measures Definition

The *angularity measure* (Bernhardt, 1992) is the supplement of the angle at each vertex (Figure 7). A set of measures based on the angularity is the *average angularity*, the *average positive angularity*, the *average negative angularity*, the *average magnitude angularity*, the *average angularity length*, the *average magnitude angularity length* and their standard deviations. *Curvilinearity* measures the number of infection points, which are the points of the line where angularity changes sign. A set of measures based on the curvilinearity are: the average *curvilinearity length*, the *standard deviation of curvilinearity length* and the *curvilinearity ratio*. The *average angularity* and *curvilinearity* can be measured at different levels of detail by computing the measures for different sets of vertices, other than the immediately consecutive ones. For every triplet of vertices n points apart, where n =1 to 60, an angularity value is computed. The average of all these values is extracted for each n and the *average angularity plot* at different ranges is constructed. In the same way, the *average magnitude angularity plot* at different ranges and the *curvilinearity plot* at different ranges are constructed. According to Jasinski (1990) the *average angularity* variable is calculated as follows: The measure can take values ranging from 0 (straight line, inter-segment angle is 0), to 1 (the line backs on itself, the angle is 180 degrees). The value of *angularity* is the same whether the angle of change is positive or negative.

Structure signature (Buttenfield, 1991) incorporates five parameters whose measurements are repeated for different resolution (Figure 8). The first computation (*anchor line length*) measures the Euclidean distance between the starting and ending points of the string. It is used to compute the

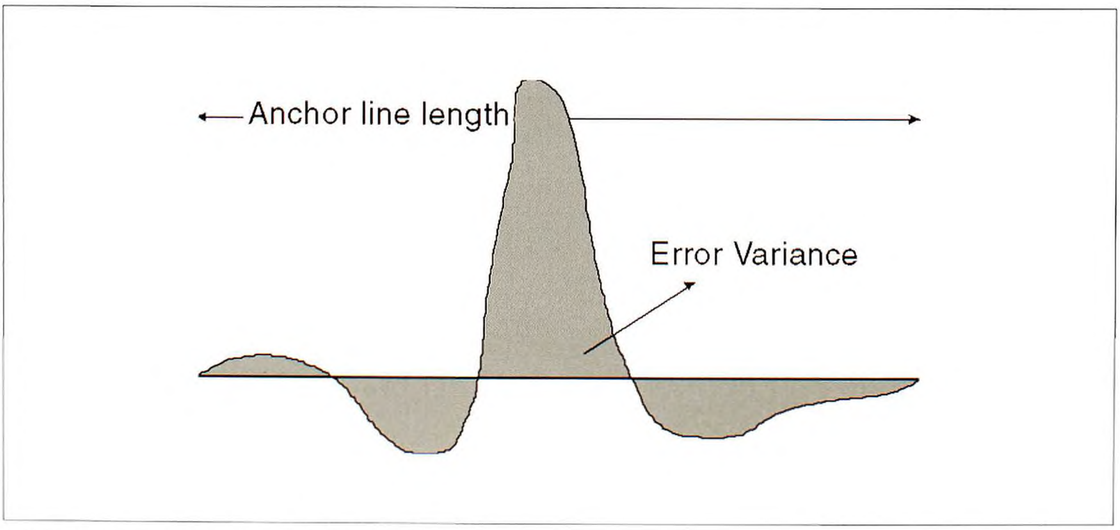


Figure 8: Parameters defining structure signature.

two parameters describing the Minimum Bounding Rectangle (MBR) of each segment.

The first one, labeled *bandwidth*, measures the maximum perpendicular deviation of any coordinate in the original string on either side of the anchor line. The second is called *segmentation* and is the distance from the coordinates of the starting point to the location on the anchor line where the maximum deviation occurs. *Error variance* is the discrete approximation of the total discrepancy between the anchor line and the original coordinate string. *Concurrence* is a count of the number of times the line segment crosses the anchor line.

Fractal dimension relates the decrease of stride length used to measure the line length, to the increase in the total line length. The plot of the set of measured lengths and strides in logarithmic scale constitutes the Richardson plot. A linear regression model is fitted to the data. The aspect of the best-fit line and the goodness of fit (R^2) are stored for each line as the fractal dimension results. Fractal dimension calculation is usually carried out with the structured walk algorithm (Bernhardt, 1992). The *ratio of the line length to the base line length* describes the deviation of the line from its simplest form, that is, the base line.

References

Antoniou, V. 'Evaluation of XML technologies' capabilities for map composition and display'. Dipl. Thesis, Cartography Laboratory, National Technical University of Athens, 2004

Bernhardt, M.C. 'Quantitative characterisation of cartographic lines for generalisation'. Report No. 425, Department of Geodetic Science and Surveying, The Ohio State University, Columbus, Ohio, 1992

Burggraf, D. 'GML and Hydrography Electronic Navigational Charts (S-57 ENC)', GML Days 2004, http:/www.gmldays.com/gml2004/presentations/GML-S57-DavidBurggraf.pdf, 2004

Burggraf, D. 'S-57 Schema and Related Tools Manual S-57/GML'. Project TR2004-250-02 – V0.1, 14-07-2004, Prepared for the UKHO, http://www.ukho.gov.uk/attachments/b2b-GML/S57Schema Manual.pdf, 2004

Buttenfield B. 'A rule for describing line feature geometry'. In B. Buttenfield and R. McMaster, (Eds) Map generalisation: Making rules for knowledge representation, Hallow, Essex, U.K.: Longman Scientific, 1991, pp 150-171

Carstensen, L. W. 'Angularity and Capture of the Cartographic Line During Digital Data Entry'. Cartography and Geographic Information Systems, 1990,17(3) : 209-24

Cecconi, A. 'Integration of Cartographic Generalisation and Multi-Scale Databases for Enhanced Web Mapping'. Ph.D. diss., University Zurich, Switzerland, 2003

DuCharme, B. XSLT Quickly , Manning Editions, (2002)

Dutton, G. 'Scale, sinuosity and point selection in digital lines generalisation'. Cartography and Geographic Information Science, 1999, 26(1), pp. 33-53

European Committee of Standardization, 1996-97, TC287, Geographic Information – Data Description

GEMURE - Generalisation and Multiple Representation for On-Demand Map Production and Delivery, 2004, http://sirs.scg.ulaval.ca/gemure/en/home.asp

Geography Markup Language (GML), Version: 3.1.0, 2004. http://www.opengeospatial.org/specs/? page =recommendation

GiMoDig - Geospatial info-mobility service by real-time data-integration and generalisation, 2004, http://gimodig.fgi.fi/

Hair, J. F., Anderson, R. E., Tatham, R. L., Black, W. C. 'Factor Analysis in Multivariate data analysis', 1998, Prentice – Hall International, pp. 87-140

Hardy, P. Multi-Scale Database Generalisation for Topographic Mapping, Hydrography and Web-Mapping Using Active Object Techniques, 2000, IAPRS, Vol. XXXIII, Amsterdam

IHO Transfer Standard for Digital Hydrographic Data. Publication S-57, Edition 3.1, November 2000

Jasinski, M. J. 'The comparison of complexity measures for cartographic lines', NCGIA Report 90-1, NCGIA, Santa Barbara, CA., 1990

Johansson M., Lars, H. 'Using Java Topology Suite for Real-time Data Generalisation and Integration', ISPRS Workshop, Ottawa, Canada, 2002

Jones, C. B., Kidner, D. B., Luo, L. Q., Bundy, G. L., and Ware, J. M. 'Database, Design for a Multi-Scale Spatial Information System'. International Journal of Geographic Information Systems, 10 (8), 1996. pp. 901-920

Mandelbrot, B. 'How long is the coast of Britain? Statistical self similarity and fractional dimension'. Science, 1967, vol. 156, pp.636-638

McMaster, R. B. 'A statistical analysis of mathematical measures for linear simplification'. The American Cartographer, 13(2), 1986, pp.103-116

Moktharian, F., Mackworth, A. K. 'Scale-based description and recognition of planar curves and two dimensional curves'. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, 8:34-43

National Oceanic & Atmospheric Administration (NOAA) Medium Resolution Digital Vector Shoreline, http://sposerver.nos.noaa.gov/projects/shoreline/shoreline.html

OGS, The OpenGIS Service Architecture (AS Topic 12) 4.3 02-112 2001-09-14 Same as ISO 19119 ISO

Open Geospatial Consortium, 2004. http://www.opengeospatial.org/

Peng, Z.-R., Zhang, C. 'The roles of GML, SVG, and WFS specifications in the development of Internet GIS'. Journal of Geographical Systems, 2004, Vol. 6 pp.95-116

Plazanet, C., Affholder, J. G., Fritsh, E. 'The importance of geometric modeling in linear feature generalisation'. Cartography and Geographic Information Systems, 1995, 22(4), pp. 291-305

Plazanet, C. 'Enrichissement des bases de donees geographiques: analyse de la geometrie des objects lineaires pour la cartographiques (application aux routes)'. These de doctorat, 1996, Universite de Marne La Valle

Scalable Vector Graphics (SVG) 1.2, W3C Working Draft 27 October 2004. http://www.w3.org/TR/SVG12/

Simple Object Access Protocol (SOAP), 2003, http://www.w3.org/2000/xp/Group/

Skopeliti A. 'Error Assessment in Cartographic Data Bases'. Ph.D. diss., 2001, Cartography Laboratory, National Technical University of Athens

Skopeliti, A., Tsoulos L. 'On the Parametric Description of the Shape of the Cartographic Line', Cartographica 1999, 36 (3), pp. 57-69

Skopeliti, A., Tsoulos, L. 'A knowledge - based approach for the cartographic generalisation of linear features', Proceedings of the 20th International Cartographic Conference, Beijing, China, August 6-10, 2001, vol. 3, pp. 1903 – 1913

Skopeliti A., Tsoulos L. 'Methodology for the assessment of generalisation accuracy'. Proceedings of the 4th ICA Workshop on Progress in Automated Map Generalisation, Beijing, China, 2001

Spanaki M. 'XML Technologies and Map Composition', MSc Thesis, 2002, Cartography Laboratory, National Technical University of Athens

Spanaki M., Antoniou B., Tsoulos L. 'Web Mapping and XML Technologies – A Close Relationship', AGILE 2004 - 7th AGILE Conference on Geographic Information Science, April 29th - May 1st, Heraklion, Greece

Thapa, K. 'Data compression and critical points detection using normalised symmetric scattered matrix', 1989, Auto-Carto 9, pp.78-89

Tsoulos, L., Spanaki M., Skopeliti A. 'An XML-based approach for the composition of maps and charts', 21st International Cartographic Conference, Durban, Africa, August 6-10, 2003

Tzamakou, T., 'Parametric description of cartographic line shape', Dipl. Thesis, 2004, Cartography Laboratory, National Technical University of Athens

Wang, Z., Müller, J-C. 'Line generalisation based on an analysis of shape characteristics', Cartography and GIS, 1998, 25 (1), pp. 3 – 15

Watt A., Designing SVG web graphics, New Riders Publishing, (2002) Web Feature Service (WFS), Version: 1.0, Document: 02-058, Date: 2002-05-17, http://www.opengeospatial.org/specs/

Web Map Service (WMS), Version: 1.01.3 Document:04-024, Date: 2004-08-02, http://www.opengeospatial.org/specs/

XSL Transformations (XSLT), Version 1.0 W3C Recommendation 16 November 1999, http://www.w3.org/TR/1999/REC-xslt-19991116

Biography of the Authors

Dr Andriani Skopeliti holds a Dipl.Eng. (1994) in Surveying Engineering from the National Technical University of Athens, Greece and a Ph.D. (2001) in Cartography from the National Technical University of Athens, Greece.
Since 1994, she has been involved in several research projects funded mostly by the EU. She teaches undergraduate and postgraduate courses in Cartography and Geographic Information Systems at the National Technical University of Athens and the University of Aegean. Her research interests include Cartography, Geographic Information Systems and Web technology.

Dr Lysandros Tsoulos graduated from the Department of Rural and Surveying Engineering (University of Thessalonica) in 1972. He received his Dr Eng. degree from the National Technical University of Athens (NTUA) in 1989. He worked at the Hellenic Navy Hydrographic Service for 17 years as head of the directorate of Cartography and the Computing Center. Since 1991 he is a member of the faculty of the NTUA.
Currently he is an Associate Professor at the School of Rural and Surveying Engineering, Vice-Chairman of the School and Director of the NTUA Geomatics Center. He has eighty publications in international/national journals and conference proceedings on Digital Mapping and Geographical Information Systems. He has participated in a number of projects like: Impact II, STRIDE, MEDSPA, WERATLAS, GEOMED and STATLAS. His research interests include: cartographic generalisation, expert systems in cartography, digital atlases and Web Technology. He is member of a number of National and International Scientific Committees.

E-mail: askop@central.ntua.gr
lysandro@central.ntua.gr