Towards a Geometrical Spatial Integration Model for GIS

ISMAIL WADEMBERE & PATRICK J. OGAO*
Department of Information Systems,
Faculty of Computing & Information Technology, Makerere University

---

Abstract. Geospatial information systems provide avenues to fulfill users' quest for both absolute and relative locations of features/objects. To achieve this, the users need to have access to different geospatial data sets from various sources. This calls for integration of data from different geospatial sources. But, there are shortcomings as different data sets may not map exactly onto one another, one of reasons being the difference in features' geometry. This paper presents a conceptual model for geospatial data integration that can identify and measure differences and adjust spatial geometries of geospatial features to form meaningful objects which can be used for geo-spatial analysis, modeling, and easy geo-information management.

Keywords: Spatial Data, Integration, Geometry Adjustment, and Information Management

---

IJCIR Reference Format: Ismail Wadembere & Patrick J. Ogao. Towards a Geometrical Spatial Integration Model for GIS. International Journal of Computing and ICT Research, Vol. 2, No. 2, pp.37-48. http://www.ijcir.org/volume2-number2/article 4.pdf.

INTRODUCTION

Geospatial information systems especially Geographical Information Systems (GIS) have enabled users to analyze, model, and visualize geospatial (space and earth related) data. This has made GIS a useful media for exchange of geo-spatial information among individuals and organizations and a basis (both as science and tool) for space and location based decision-making.

However, it takes a lot of time, effort, resources, and skills to create and store geo-spatial data and because of that, in most situations many practitioners do not have all the data they need in one database (FGDC, 2007). To meet this need, many individuals and organizations use geo-spatial data from different sources. This has made geo-spatial data integration difficult to ignore as there is need to take advantage of different spatial data in different locations to reduce on the cost and time involved in data production.

Geospatial data integration is not straight forward as these data sets are collected by many individuals/organizations and sometimes using different methods, standards, data models, and information technologies. As result, the data may vary in spatial geometries causing heterogeneity, which can be zonal/horizontal or layered/vertical fragmentation (Kampshoff, 2006). In this paper, we handle zonal heterogeneity which is as a result of geometrical difference in thematically similar data sets that can be evidenced by incomplete or inaccurate mapping of the data sets.

---

* Author's Address: Ismail Wadembere*, Patrick J. Ogao; Faculty of Computing & Information Technology, Makerere University P.O. Box 7062, Kampala, Uganda, iwadembere@cit.mak.ac.ug, ogao@cit.mak.ac.ug

International Journal of Computing and ICT Research, ISSN 1818-1139 (Print), ISSN 1996-1065 (Online), Vol.2, No.2 pp. 37-48, December 2008.

This paper starts by looking at the need for geospatial data integration which motivates investigation into the different data integration approaches, spatial representation, geometrical integration and their limitations. We use the outcomes of that research to triangulate different methods as we move towards designing specifications for a Geometrical Spatial Integration Model (GSIM). We end this paper by a look at proposed future work and conclusion.

2. THE NEED FOR GEO-SPATIAL DATA INTEGRATION

Currently, we are seeing developments in technologies which are creating services which need geospatial data comparison (Najar et al., 2006) and integration (Sester et al., 2007). This is evident with increased geo-spatial developments like Google Earth, Microsoft's MapPoint.Net, Oracle Locator and Spatial, and ESRI's ArcGIS; new platforms, vendors, users, and conferences and publications on location-based services (Zhao and Motorola, 2002), (Chang et al., 2007), (Sester et al., 2007), (Busgeeth and Rivett, 2004). These developments and changes although seemingly diverse, are related as they all use and take advantage of the geospatial information available (Sagayaraj et al., 2006). That is why we are seeing increased importance of location in information systems and services (Strader et al., 2004).

Increasingly, technologies like the Internet, web services, image capture and communication, low cost spatial data collection tools, grid computing power, and On-Line Transaction Processing (OLTP) application (Skog, 1998), (Sharma, 2005), (Sohir, 2005) among others utilize geo-spatial data. In the process, technologies coupled with the above trends are positively influencing the need for spatial data integration.

The need for geospatial integration is further fuelled by Location Based Services and Location-Tracking trends; for example in the United States the move to location-aware mobile phones have been driven by the federal government's enhanced 911(E911) mandate (FCC, 2006). Also, researchers are experimenting with ideas such as leaving messages for other people attached to certain places, all enabled by on-line location services (Sharpe, 2002). All these examples underline the emergent need for geospatial data integration to improve on sharing and information management (Shi and Yeh, 1999).

Further use of geospatial data and need for integration can be seen in Enterprise Location Services and Software (ELSS); which concern the use of location-specific information within information systems. Several technologies combine to make ELSS including the Service-Oriented Architectures (SOA) and the automated location determination technologies like GPS, cellular networks, and radio frequency. Also, spatially-enabled IT Infrastructure, Open Standards like Geographical Markup language (GML), Spatial Data infrastructures (SDI), 3G System (Zhao and Motorola, 2002), etc are becoming very common (Sharma, 2005, Sohir, 2005).

All these technologies provide avenues for sharing geo-information from different sources. But for many organizations, integration is still a major challenge in GIS projects. The need for integration arise from the nature of geo-spatial data and information - huge data sets, different formats, and the need to share information between organizations and individuals (Evans, 1997), (Erdi and Sava, 2005). However, most data sources, databases and information systems are not designed for integration. Thus, whenever integrated access to different source systems is desired, the sources and their data that do not fit together have to be coalesced using additional adaptation and reconciliation functionality. The goal is always to provide a homogeneous, unified view on data from different sources.

Integration tasks may depend on various factors such as those (Ziegler and Dittrich, 2004) list: “the architectural view of system; the content and functionality of the component systems; the kind of information that is managed by component systems (spatial data, alphanumeric data, multimedia data; structured, semi-structured, unstructured data); requirements concerning autonomy of component systems; intended use of the integrated system (read-only or write access); performance requirements; the available resources (time, money, human resources, know-how, etc.); and the source and type of method used to collect the data”.

Additionally, several kinds of heterogeneity typically have to be considered. These include differences in (i) hardware and operating systems, (ii) data management software, (iii) data models, schemas, and data semantics, (iv) middleware, (v) user interfaces, and (vi) business rules and integrity constraints (Ziegler and Dittrich, 2004). For geometrical spatial, we have mainly zonal/horizontal and layered/vertical heterogeneity.

Layer fragmentation occurs when a base-layer on which the integrated multi-layer datasets depend is replaced by an updated version but the rest of the new layers are collected using the old one. This is termed as systematic tessellation (Masuyama, 2006) which is the difference due to use of different geodetic models when capturing layers. Zonal heterogeneity is the result of capturing thematically similar data independently by different individuals with no common framework being followed. Examples of problems include: - same points with varying coordinates, arcs which are supposed to be the same intersecting, neighboring areas overlapping each other, areas which are supposed to be touching each other leaving opening, etc (see Figure 3 below). These lead to geometrical and topological inconsistencies like creating sliver polygons, dangling arcs and nodes. It is zonal heterogeneity which is the focus in this paper.

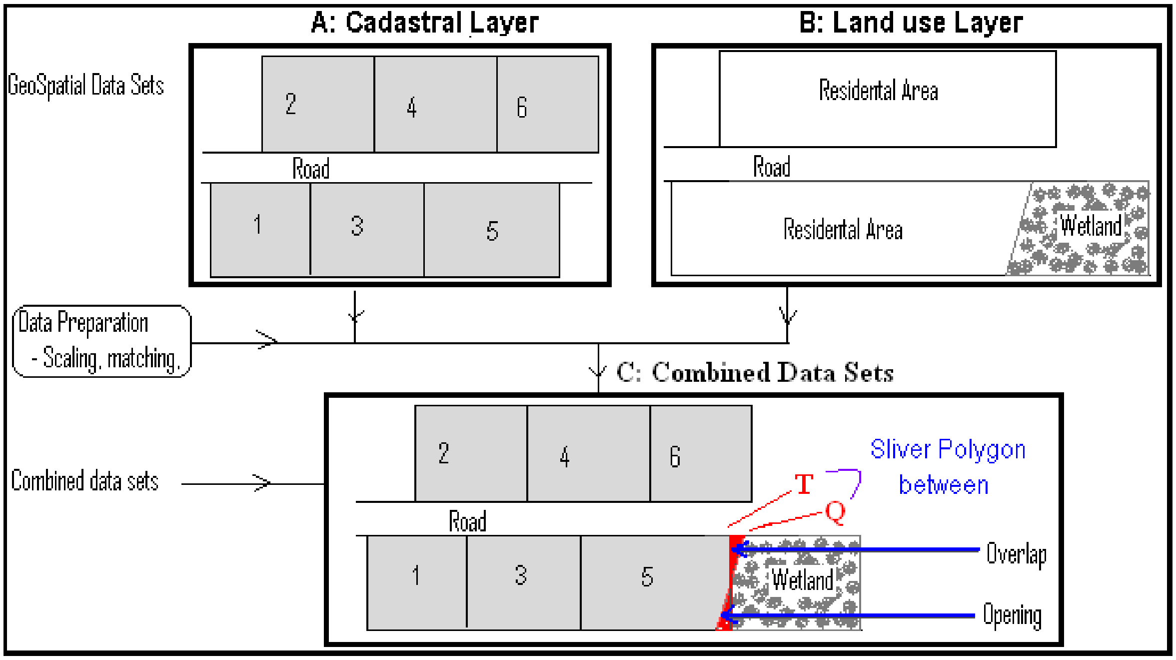

Figure 3: Geometrical Integration Problem

Some of the factors which cause geometry differences are presented on the Figure 3 above which shows two layers of vector data. Layer A: cadastral layer having plots 1, 2, 3, 4, 5, and 6 demarcated for construction of residential houses and layer B: land use layer which indicates the areas set aside for different land use activities (residential use and wetlands). When we carry out data preparations (including scaling and matching) on the two layers and we combine them by merging their geometries and observe the result on layer C (combined data sets). We notice that plot 5 has an overlap with the wetland and also there an opening which is neither plot 5 nor a wetland. This is when we need to adjust the geometries of individual features/objects so that we eliminate the overlaps and openings which create sliver polygons (like polygon between points P & Q), dangling arcs, and nodes.

3. DATA INTEGRATION APPROACHES

Different researchers ((Ziegler and Dittrich, 2004), (Nebert, 2003), (Sohir, 2005), (Musinguzi et al., 2004), (Evans, 1997), (Nebert, 2003), (Erdi and Sava, 2005), (Kilpelainen, 1997), (Friis-Christensen et al., 2005), (Skog, 1998), (Sharma, 2005), (Sester et al., 2007), (Bell, 2005), and so on) present different ways of addressing the integration problem which can be done at presented system layers, data storage, open systems, and setting sharing channels. These include:

a) *Manual Integration:* users directly interact with all relevant information systems and manually integrate selected data. The users have to deal with different user interfaces and query languages that require detailed knowledge on location of data, logical data representation, and data semantics.

b) *Peer-to-peer (P2P) integration*: a decentralized approach to integration between distributed and autonomous peers where data can be mutually shared and integrated. P2P integration constitutes, depending on the provided integration functionality, either a uniform data access approach or a data access interface for subsequent manual or application-based integration.

c) *Common User Interface*: user is supplied with a common user interface (e.g. web browser) that provides a uniform look and feel when accessing data from different relevant information systems. System are separated but homogenization and integration of data is done by the tools like search engines.
d) *Integration by Applications*: integration applications access various data sources and return integrated results to the user. This is mostly suitable for systems with a few components since applications become many as the number of system interfaces and data formats to homogenize and integrate grows. Examples include workflow management systems that allow implementation of business processes where each single step is executed by a different application.
e) *Integration by Middleware*: Middleware provides reusable functionality that is generally used to solve dedicated aspects of the integration problem e.g. SQL-middleware. Applications are needed in the form of middleware tools and are usually combined to build integrated systems and they are relieved from implementing common integration functionality.
f) *Uniform Data Access*: logical integration of data is accomplished at the data access level. Global applications are provided with a unified global view of physically distributed data, though only virtual data is available on this level. However, global provision of physically integrated data can be time-consuming since data access, homogenization, and integration have to be done at runtime. Examples include mediated query systems, portals for Internet/intranet, federated database systems, etc.
g) *Open systems*: geospatial data integration is done using distributed models that are based on open systems technologies like OpenGIS, XML, GML, etc.
h) *Geo-information sharing infrastructures*: communities agree on how to share data like SDI, Geospatial Data Clearinghouse, Proprietary protocols, GIS-network integration, CRUD (Create, Read, Update, and Delete) matrix where modeling tool is used in data-oriented methodologies to show processes and their associated data. CRUD is a spatial data integration approach based on comparing organizations spatial data in a matrix. It shows data sets on the vertical axis and their associated departments or users on the horizontal axis. The matrix provides information on data ownership, maintenance, use, and geographical distribution.
i) *Common Data Storage*: Here, physical data integration is performed by transferring data to new data storage; local sources can either be retired or remain operational. This approach provides fast data access. Examples include federated database systems, data warehouses (data from several operational sources like OLTP are extracted, transformed, and loaded (ETL)), storing geo-data with varying standards in same geodatabase using multiple representation, and attribute versioning.

However, these approaches do not meet all the requirements especially zonal heterogeneity, as some times, there is need to change and adjust geometries of spatial features (Peng, 2005), so that data with varying geometries can fit exactly on each other to allow the process of integration to succeed. Thus, in this paper we base our approach on philosophy of common data storage to develop a conceptual GSIM which can handle geometries to solve zonal heterogeneity. The model is to support the determination of differences between the geometries of geo-spatial datasets; then to adjust these geometries so that data can be integrated before they are adapted/adopted in sharing avenues and used in spatial analysis and modeling. Before that, let us look at spatial data representation and geometrical integration to put our approach into perspective.

4. SPATIAL DATA REPRESENTATION, ONTOLOGY, AND GEOMETRICAL INTEGRATION

Geo-Spatial data is data that occupies cartographic (map-able) space that usually has specific location according to some geographic referencing system (e.g. x-y) or address. This data is represented using two major data structure: raster and vector, which help to store spatial characteristics of features. Vector data are based on features and have x and y coordinates. Raster data are based on pixels, with a grid like system of rows and columns (Burrough, 1986), (Nievergelt and Widmayer, 2006).

In developing GSIM, our focus is vector data. Vectors are scalable drawings created from mathematical calculations and stored in the computer as mathematically defined geometrical feature primitives (points, lines/curves, and polygons) making them easy to create and edit. Vectors can be reduced and enlarged with little loss of quality because the equations can easily be recalculated at new scales (Chrisman, 1997).

A geometrical point is a one-dimensional feature (has length only and no width). It is a spot (location) that has no physical or actual spatial dimension. Points indicate specific location of features, which are usually not shown in true size, especially for things that are too small to depict properly at a given scale. Points also show location of intangible (non physical entities) features e.g. address or location of occurrence like traffic accidents. A lines has a beginning and an end and are linear features; either real (e.g. road) or administrative (e.g. boundary). Sometimes the thickness of the line indicates a measure (such as amount of traffic on a road) or type of road (such as primary verses secondary). A polygon is an enclosed area, a two dimensional feature with at least 3 sides (and therefore with an inside area and perimeter) (Burrough, 1986).e.g. plot of land.

The geometrical feature primitives combine to form spatial objects which can be grouped into layers, also called overlays, coverages, or themes. One layer may represent a single entity type or a group of conceptually related entity types e.g. a layer may have only river or may have rivers, lakes, and swamps. Options depend on the system as well as the database model and some spatial databases have been built by combining all entities into one layer (Goodchild, 1992).

Elements of reality modeled in a GIS database have identities (Tomlin, 1990):

(i) The element in reality – entity "a phenomenon of interest in reality that is not further subdivided into phenomena of the same kind",
(ii) The element as it is represented in the database – object "a digital representation of all or part of an entity",
(iii) The symbol that is used to depict the object/entity as a feature on a map or other graphic/cartographic display.

Geometrical integration incorporates schema integration and solving semantic conflicts of the datasets (Kampshoff, 2005). The interest here is the schema integration at the model level as a spatial feature can be broken into primitives (point, line/arc, and polygon). The primitives are expected to be the same for similar thematic datasets (datasets representing the same feature), but it is not always the case. In some situations, we have spatial geometrical heterogeneity existing which can be zonal/horizontal or layered/vertical fragmentation (Kampshoff, 2005) as previously explained.

In this research, we look at zonal heterogeneity, where the differences have to be determined and adjusted in order to have physical data integration at layer levels. (Casado, 2006), (Lee et al., 2002), and (Kampshoff, 2005) present several approaches to geometrical integration which can not be directly applied to GSIM as explained below:-

a) Similarity or Helmert Transformation deals with whole layer being treated as solid image where any geometry adjustment which is done changes the whole layer; it preserves the angles between features but not the distances. This approach is not applicable in our situation as we do not consider a layer as one solid object but we are interested in changing the geometry of a single element/facture/primitive on the layer.
b) Affine Transformations preserves neither distances nor angles although it can change the geometrical of the whole layer. This approach can not accomplish our need of just adjusting specific feature on the layer.

c) Rubber sheeting method allows local distortions and preserves topological properties by distributing random errors through out the data set. This would be good if the aim was to overcome the problem of having other feature appearing more changed than others. It is also not an individual feature change method and it might bring more problems as it would distribute errors to other features without errors.
d) Complex Conformal Mapping preserves angles and multiplies the length of short segments by a certain factor as it tries to keep the features on the layer undistorted. This approach is for whole layers and may not be applied to a single feature of on a layer.
e) Other geometric correction algorithms like one developed by (Lee et al., 2002) are for satellite image correction as the parameters considered in the correction are instrument characteristics, the

satellite motion, earth rotation, terrain effects, and the ground control points (GCP) which can not be directly applied in vector GIS data.

f) Stochastical geometrical integration tries to balance the geometrical error available in the layer and in the process does not reproduce the values for data points as the random error is filtered in these locations.
g) Deterministic geometrical integration models can be applied when the random error is of negligible magnitude. That means that if the separation between features is very big, this approach can not be applied.
h) Kampshoff (2005) combines the last two approaches to come up with an improved geometrical integration model. Since this approach needs several points from a true map which have known coordinates in order to carry out adjustment, it can not be applied in our situation where we may need only to move a feature without using known points as they may not be available.

5. TRIANGULATION OF METHODS FOR GSIM

With no adequate method to handle geometrical zonal heterogeneity, we need to develop one but relaying on existing science approaches. For established fields of science like Information Systems (IS), academicians often take different approaches in order to solve different problems, and expand on the researches. There are many reasons for this including (King and Lyytinen, 2005), (Benbasat and Zmud, 1999), (Niehaves, 2005), (Wangler and Backlund, 2005), (Erdi and Sava, 2005), (Aronson, 1998), (Bernard et al., 2003), (Mingers, 2001) : -

a) The need to achieve a relation between IS and other fields.
b) Different approaches have been used by researchers as it gives them the opportunity to be creative.
c) The question of whether existing research approaches provide a suitable basis to start new research.
d) New researchers in IS with little or no knowledge about IS research approaches and paradigms end up creating new approaches.
e) The rapid change in Information Technology (IT) has led IS researchers to adopt or create new approaches which can cope with IT changes.
f) The changing needs of users such as the use of geo-spatial/location data in many IS
g) The emergence of open standards has forced researchers to change approaches.

Diverse academic disciplines, approaches, and research communities contribute to IS research (Niehaves, 2005). It is against this background of diversity that we develop GSIM using multi-methods so that it can benefit from such a variety of approaches and contributions. In the process, it helps us to see IS-related phenomena from different points of view (Mingers, 2001), (Weber, 2004), thus eventually benefiting the GSIM development. Among the issues that we considered was to develop GSIM using different approaches/paradigms basing on independent components/units/object/sub-models, which can function both as stand alone or integrated with others to accomplish a task.

This means that GSIM is not tied to one format (Bernard et al., 2003) and changes are implemented by modifying specific component concerned. Emphasis will be placed on developing the GSIM that can be implemented using messaging, UML, GML, and XML manipulation for identifying data components and trigging different units/objects. The development of the GSIM will also utilize spatial validation, spatial matching, spatial interpretation, reverse engineering, and system thinking (Aronson, 1998), and GIS spatial functions such as overlay, merge, buffer, connectivity, neighborhood, adjacency, and network operations.

6. SPECIFICATIONS AND DESIGN OF GSIM

Let us assume a user/practitioner has two geo-spatial data sets from different sources and s/he is interested in using the two data sets, thus s/he has to integrate/merge the two data sets. Several integration issues and different questions emerge which need to be considered and solved; these formed the basis for specification and design of GSIM

Integration issues considered in the specification and design of GSIM: -

a) geo-spatial data may contain bias like some features being over emphasized and some under represented
b) different data sets representing the same entity may extend over a different area
c) data sets may just be incomplete i.e. do not have all the required elements for certain analysis and modeling to be complete or/and meaningful
d) sources may contain closely related and overlapping data
e) data may be stored in multiple data models and schemas
f) topology requirements like overlap, containment, connections, and adjacency may be missing or not proper
g) data sources may be having differing query processing capabilities

Questions asked in order to define a problem that formed the basis for specification and designing GSIM: -

i) When are two spatial data sets different?
ii) Is the difference between data sets due to geometry?
iii) How much of the difference is due to geometrical difference?
iv) Which of the data sets is correct or better?
v) What parameters of geometry are different?
vi) Can the geometrical parameter be changed?
vii) How can geometrical parameters be adjusted?
viii) How can a combined spatial geometry be built?
ix) How can data sets be combined?

With those integration issues (a-g) and questions (i-ix) in place, we converted them into standard spatial meaning to get the specific components of GSIM which are depicted on Figure 4 below and explained in the sections following it; but the conversion and filtering was done as follows:

- *Spatial scaling* handles and is got from integration issues a & b and questions i & ii
- *Spatial data matching* handles and is got from integration issues a & b and questions i, ii, & iii
- *Best spatial data set* handles and is got from integration issue c and question iv
- *Spatial components identification* handles and is got from integration issue d and question v
- *Geometrical spatial difference* handles and is got from integration issues d & e and question vi
- *Geometrical spatial adjustment* handles and is got from issue e and question vii
- *Geometrical spatial builder* handles and is got from question viii
- *Topology creation* handles and is got from integration issue f
- *Final model development* handles and is got from integration issue g and question ix

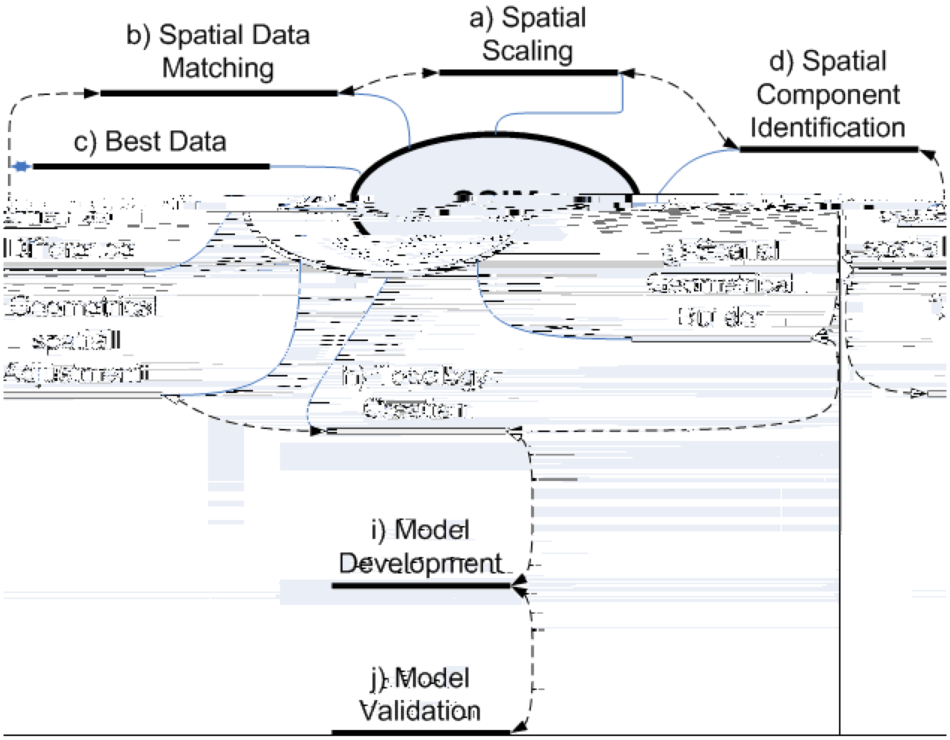

Figure 4: Geometrical Spatial Integration Model Components

Spatial Scaling

Two data sets are checked to see if they occupy the same aerial extent, if not the data sets are changed so that they occupy the same area. This may involve a change of scale because of shrinkage or enlargement of a particular data set. If data sets representing the same area have different scales, then scale change has to be done so that the data sets cover the same area extent. The aim here is not to handle whole scale of the data within the object although scale change can take place so that the datasets cover the same area extent. The scale change, which is handled by the GSIM, affects individual geometries of the spatial elements for example changing the size of polygon say a plot of land in a cadastral layer.

Spatial Data Matching

The main purpose for this object is to determine if two spatial data sets are different and if the difference is due to geometry. Here, we also determine which geometrical parameters are causing the difference. This involves spatial object matching to predict errors due to inexact matching, which end up producing for example sliver lines/polygons or opening.

We employ the ideas of spatial data matching and spatial interpretation in determining the spatial data sets with similar geographies and comparing spatial data sets at object level; we take advantage of the computational model developed by (Masuyama, 2006) . The end product at this level is an algorithm that can determine spatial geometry differences and identify which components of the geometry are causing the mismatch. The main considerations are scale, resolution, and actual errors in a data set.

Best Spatial Data Set

This compares different data sets using known data sets (especially base/datum data set) or known controls. The purpose is to identify correct data sets or to assess which data sets should be used as the basis and whose integration parameters will be used during adjustment.

Spatial Component Identification

As the various components of geo-spatial data sets have to be handled differently during special adjustment, the role of this component is to breakdown the different primary component of the data set into individual layers i.e. point layer, ployline layer, and polygon layer.

Geometrical Spatial Difference

This is used for geometrical quantitative difference determination. This involves the process of extracting the geo-data's spatial geometry and determining the geometry relationships of components depending on the particular geo-data source.

Spatial validation is utilized to determine elements and is based on a) Logical cartographic consistency such as overlap, containment, adjacency, and connectivity, b) Closed polygons, c) One label for each polygon, d) No duplicate arcs and e) No overshoot arcs (dangles). GIS spatial functions like overlay, merge, connectivity, neighborhood, operations are also utilized to develop an algorithm for geometrical quantitative spatial difference determination. Current computational models which are based on geometric features like one suggested by (Masuyama, 2006) are be employed.

The outcome is to be able to handle:-

- *For line features*, the whole line feature is picked at a time and comparison takes place in the following parameters:- length of the ploylines, location (using coordinates) of the nodes and vertices, overlap, number of nodes, thickness, and continuity (connectivity).
- *For the point features*, an imaginary fine grid is over laid on the two data sets so that the values in each grid can be compared. The grid size is according to the scale of the data set i.e. the grid square is determined by the scale at which the comparison is being done. The difference is determined and comparison in precision, resolution, and actual data values is done. If the difference is zero, we assume that the two cells are the same and if the resultant is not zero (could be positive or negative), then the two are different.
- *For polygons*, whole polygons are compared, the shape, corner/edge coordinates, overlaps, label points, adjacency, containment, over size, under size, neighborhood, not closed, and does not have a label.

Geospatial Ontology and Geometrical Spatial Adjustment

After identifying the components causing the differences and determining how much of the difference exist; this unit/object breaks down geometries of data to individual primitives (points, arcs and polygons). This is done as geometrical differences (not semantics differences as described by (Budak Arpinar et al., 2006)) within individual features among thematically similar data sets can still remain present even after overlaying and merging. So the adjustment should be done on the primary elements – the primitives of individual features.

This approach is being adopted as spatial data has been conceptually defined as geometric data consisting of points, arcs and polygons (Lu et al., 2007) and these conceptual objects (geometric data) are used to represent GIS objects/features which should be robust and efficient in terms of storage, retrieval, processing and good graphical reconstructing.

However, to handle the different primitives at the same time brings in challenges in terms of requiring different algorithms to adjust the different primitives (Masuyama, 2006). This is so as it can be observed by taking a close look at the different primitives. It is evident they have different geometries and situation could be made simpler if only one type of primitive can be handled during geometry adjustment. In this research we look at disintegrating the lines/arcs and polygons into points as a move towards providing only one primitive type (points) on which geometry adjustment can take place.

Spatial validation and Spatial Reverse Engineering (SRE) are employed here to help in disintegrating geo-data from any source into primitive objects/elements. This is done by considering the strength and goals of SRE (i) to analyze a data set, (ii) to identify the data's components, (iii) to identify the interrelationships of the data components, and (iv) to create representations of the data in another form or at a higher level of abstraction. SRE helps to return to a less complex or more primitive state or stage of the spatial data geometry. The resultant algorithm has to accomplish geometrical point correction according to the geometrical spatial difference of features which could have comprised of ploylines, points and polygons.

Geometrical Spatial Builder

This object looks at the different spatial primitives that have been adjusted and builds them into features which have spatial meaning; these are later combined during topology creation and building.

Topology Creation and Building

As GIS data should have proper topology to enable users to carry out modeling and analysis, this component combines the spatial objects generated using geometrical spatial builder into proper polygons, ploylines, and points and creates relationships between them to obtain meaningful spatial data sets.

Final Model Development

This is the combining of data sets after the above components/objects have played their roles. After developing the different components, the resultant GSIM is achieved using an iterative process where the design rules for each possible sequence of the object are taken into consideration and deployed according to given user needs/requirements. The resultant GSIM and also the different objects can be deployed using approaches like novel service retrieval (Klein and Bernstein, 2004) that capture service semantics using process models and apply a pattern-matching algorithm to find the services with the behavior the user wants.

Model Validation and Evaluation

To make sure the model can be utilized; GSIM will be evaluated with the help of topological requirements testing. This will be carried out both theoretically and experimentally using the following criteria: the maturity and capabilities of each object, whether or how well each object can be used in conjunction with others, and how the whole model can fulfill geometrical integration requirements. A goal-based evaluation will be performed to decide if our goals are met. We will look for logical cartographic consistency involving testing for overlap, containment, adjacency and connectivity. This requires all polygons be closed, there be no duplicate arcs and no overshoot arcs (dangles) and only one label for each polygon, and so on. GIS spatial functions like overlay, merge, connectivity, neighborhood, and operations will also be utilized in testing.

7. ZONAL HETEROGENEITY AS ADDRESSED BY GSIM

Zonal heterogeneity being evidenced by geometrical variations which produce sliver polygons, dangling arcs, nodes, etc (see Figure 3) during integrating/merging thematically similar data features/layers is caused by various factors including storing data in multiple data models and schemas, capturing data independently by different individuals with no common framework being followed, etc is addressed by above steps (see Figure 4) which form components of GSIM. With these GSIM components, GIS practitioners should be able to identify the geometrical variations and eliminate them, which provide a smooth and easy approach to merging different geo-spatial data sets.

8. FUTURE WORK

The conceptual GSIM will be implemented and tested to determine if it meets the need of the geo-spatial data integration. We will examine how GSIM can take advantage of the expanding geo-technologies like GML. Subsequently we plan to investigate if it can be adapted and used in spatial analysis and modeling which can feed into sharing avenues like SDI and general go-information management.

9. CONCLUSION

The main function of geospatial information systems is to provide location specific information which is easy to understand and can be used in decision-making. In this paper, we have presented a conceptual

Geometrical Spatial Integration Model (GSIM) which when implemented can identify and compare differences and adjust geometries of primitives (points, lines, and polygons) so that they fit exactly to come up with complete and meaningful spatial features. With this model in place, geospatial practitioners may be able to identify sliver primitives and features in data sets which can be eliminated, making integration easier, and leading to increased use of different data sets created and kept by different individuals. This may facilitate natural resources managers/users access to geo-information, and lead to improved analysis and modeling results, reduced cost and time involved in geo-spatial data production, and informed location-based decision-making.

10. REFERENCES

ARONSON, D. (1998) Systems thinking. *R\&D Innovator. (now Innovative Leader),* 6.

BELL, C. A. (2005) Attribute-Level Versioning: A Relational Mechanism for Version Storage and Retrieval. *School of Engineering.* Richmond, Virginia (USA), Virginia Commonwealth University.

BUDAK ARPINAR, I., SHETH, A., RAMAKRISHNAN, C., LYNN USERY, E., AZAMI, M. & KWAN, M. P. (2006) Geospatial Ontology Development and Semantic Analytics. *Transactions in GIS,* 10**,** 551-575.

BURROUGH, P. A. (1986) *Principles of geographical information systems for land resources assessment,* Oxford, Oxford University Press.

BUSGEETH, K. & RIVETT, U. (2004) The use of a spatial information system in the management of HIV/AIDS in South Africa. *International Journal of Health Geographics* 3**,** 13.

CASADO, M. L. (2006) Some Basic Mathematical Constraints for the Geometric Conflation Problem. IN CAETANO, M. A. P., M. (Ed.) *7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences.*

CHANG, C.-L., CHANG, Y.-H., CHIANG, M., CHUANG, T.-R., DENG, D., HO, S. & LIN, F.-T. (2007) Experience in Building GML-based Interoperable Geo-spatial Systems. Institute of Information Science, Academia Sinica, Taipei, Taiwan.

CHRISMAN, N. R. (1997) *Exploring geographic information systems,* New York, John Wiley and Sons. Inc.

ERDI, A. & SAVA, D. S. (2005) Institutional Policies on Geographical Information System (GIS) Studies in Turkey. *Pharaohs to Geoinformatics, FIGWorkingWeek 2005 and GSDI-8.* Cairo, Egypt.

EVANS, D. J. (1997) Infrastructures for Sharing Geographic Information among Environmental agencies. Massachusetts Institute of Technology.

FCC (2006) Federal Communications Commission. Public Safety and Homeland Security Bureau, USA

FGDC (2007) The Federal Geographic Data Committee.

FRIIS-CHRISTENSEN, A., NYTUN, J. P., JENSEN, C. S. & SKOGAN, D. (2005) A Conceptual Schema Language for the Management of Multiple Representations of Geographic Entities. *Transactions in GIS,* 9**,** 345-380.

GOODCHILD, M. F. (1992) Geographical information science. *International Journal of Geographical Information Science,* 6**,** 31-45.

KAMPSHOFF, S. (2005) Mathematical Models for Geometrical Integration. *First International Workshop on Next Generation 3D City Models.* Bonn, Germany, Geodetic Institute, RWTH Aachen Univeristy, Germany.

KAMPSHOFF, S. (2006) Mathematical Models for Geometrical Integration. Geodetic Institute, RWTH Aachen Univeristy, Germany.

KILPELAINEN, T. (1997) Multiple Representation and Generalization of Geo-databases for Topographic Maps. Finnish Geodetic Institute.

KLEIN, M. & BERNSTEIN, A. (2004) Toward high-precision service retrieval. *Internet Computing, IEEE,* 8**,** 30-36.

LEE, Y. Y., WU, A. M. & WU, F. (2002) An Algorithm for Geometric Correction of High Resolution Image Based on Physical Modeling. *The 23rd Asian Conference on Remote Sensing.*

LU, C.-T., SANTOS, R. D., SRIPADA, L. & KOU, Y. (2007) Advances in GML for Geospatial Applications. *Geoinformatica,* 11**,** 131–157.

MASUYAMA, A. (2006) Methods for detecting apparent differences between spatial tessellations at different time points. *International Journal of Geographical Information Science,* 20**,** 633-648.

MINGERS, J. (2001) Combining IS research methods: towards a pluralist methodology. *Information Systems Research,* 12**,** 240-259.

MUSINGUZI, M., BAX, G. & TICKODRI-TOGBOA, S. S. (2004) Opportunities and Challenges for SDI development in development countries - A Case study of Uganda. *Proc. 12th Int. Conf. on Geoinformatics - Geospatial Information Research: Bridging the Pacific and Atlantic.* University of Gvle, Sweden, Geoinformatics.

NAJAR, C., RAJABIFARD, A., WILLIAMSON, I. & GIGER, C. (2006) A Framework for Comparing Spatial Data Infrastructures An Australian-Swiss Case Study. *GSDI-9 Conference Proceedings.* Santiago, Chile.

NEBERT, D. (2003) *Issues of Data Standardization*, Global Spatial Data Infrastructure (GSDI).

NIEHAVES, B. (2005) Epistemological perspectives on multi-method information system Research" , . European Research Center for Information Systems, University of Mnster, Leonardo-Campus 3, 48140 Mnster, Germany.

NIEVERGELT, J. & WIDMAYER, P. (2006) Spatial data structures: Concepts and design choices. *Handbook of Computational Geometry,* 725–764.

PENG, Z. R. (2005) A proposed framework for feature-level geospatial data sharing: a case study for transportation network data. *International Journal of Geographical Information Science,* 19**,** 459-481.

SAGAYARAJ, F., P.THAMBIDURAI, B.S.BHARADWAJ, G.N.BALAGEI & N.HEMANT (2006) An Improved and Efficient Storage Technique for GIS Geometric Primitives Based on Minimum Bounding Rectangles. *9th Annual International Conference, Map India 2006.* India.

SESTER, M., GÖSSELN, G. V. & KIELER, B. (2007) Identification and adjustment of corresponding objects in data sets of different origin. *10th AGILE International Conference on Geographic Information Science 2007.* Aalborg University, Denmark.

SHARMA, J. (2005) Spatial Database Architecture - Influence of IT Trends on Spatial Information Management. *Directions Magazine.* Directions Magazine.

SHARPE, B. (2002) Trends in Information Technology. *The Appliance Studio Ltd.*

SHI, X. & YEH, A. G. O. (1999) The Integration of Case-based Systems and GIS in Development Control. *Environment and Planning B,* 26**,** 345-364.

SKOG, J. (1998) Trends in applying spatial Technologies. *13th ESRI European User Conference.* France.

SOHIR, M. H. (2005) The Role of ESA in Building the Egyptian Spatial Data Infrastructure (ESDI) Towards the Electronic Government (E-gov.). Spatial Portals and e-Government. *From Pharaohs to Geoinformatics, FIG Working Week 2005 and GSDI-8* Cairo, Egypt.

STRADER, T., TARASEWICH, P. & NICKERSON, R. C. (2004) The State of Wireless Information Systems and Mobile Commerce Research *Information Systems and e-Business Management,* 2**,** 287-292,.

TOMLIN, C. D. (1990) *Geographic information systems and cartographic modeling*, S. l.]: Englewood Cliffs.

WANGLER, B. & BACKLUND, A. (2005) Information Systems Engineering: What is it? *CAiSE Workshops,* 2**,** 427-437.

WEBER, R. (2004) The Rhetoric of Positivism versus Interpretivism. *MIS Quarterly,* 28**,** iii-xii.

ZHAO, Y. & MOTOROLA, I. (2002) Standardization of mobile phone positioning for 3G systems. *Communications Magazine, IEEE,* 40**,** 108-116.

ZIEGLER, P. & DITTRICH, K. R. (2004) Three decades of data integration - all problems solved? *Database Technology Research Group.* Winterthurerstrasse 190, CH-8057 Zurich, Switzerland, Department of Informatics, university of Zurich.