International Journal of Emerging Engineering Research and Technology
Volume 3, Issue 10, October 2015, PP 19-36
ISSN 2349-4395 (Print) & ISSN 2349-4409 (Online)

Developing Spatial Data Infrastructure in Egypt

Amr H. Ali

Surveying Engineering Department, Benha University, Cairo, Egypt (Associate Professor)

ABSTRACT

The deficiency of the existing complex environment of different data formats, scales, storage, and accessibility with no specified user permissions increased the need of efficient and accurate digital maps. Accordingly, Egypt nowadays needs to proceed with the development of data management, as well as the analysis of data and the production of comprehensible and accurate standard spatial database container to mitigate the decision-making processes.

The main objective of this research is to originate the Egyptian Stranded Spatial Data Infrastructure (ESDI). This involves not only, the generic conceptual frames of the ESDI, but also a physical implementation of its core components. Moreover, it provides the design of the spatial database model hand in hand proposing the characteristics and specifications, as well as a framework for the data migration process from diverse heterogeneous data sources. We are going to define the quality assurance guidelines and the quality control procedures that must be applied. Furthermore, the research is going to examine and improve data holding capacity, increase data accessibility, and interoperability according to specific permissions/rights.

Keywords: Standardization, Spatial Data Infrastructure, Data Migration, Data Accuracy, Interoperability.

INTRODUCTION

Since the emergence of information and communications technology, traditional paper maps have been moved to digital maps based on GIS environment, where all types of spatial data are manipulated and analysed. Despite the fact that this digital technology enables the reuse of the spatial information; it affects the incompatibility and inconsistencies of datasets, due to the multiple copies of data. Another aspect that released a new paradigm in spatial data handling and sharing across different communities is the widespread of the internet and diffusion of the computer literacy [1].

A key characteristic of spatial data is its potential for multiple fields through correlation and integration of different data sets to provide new useful insights into the interaction of many geographic phenomena. For example, recent advances in understanding the needs and approaches to sustainable development that address the economic and environment dimension have significantly benefit from the integration of different datasets, where a necessity of this integration is that the data must be spatially referred in a consistent manner [2].

Generally, spatial data is energetic to make sound decisions at the local, regional, and global levels, where, decision-makers in multifaceted fields are benefiting from spatial data, associated with the related infrastructures. This leads to the Spatial Data Infrastructure (SDI) that support information discovery, access, and the use of this information in the decision-making process [3]. Furthermore, it encapsulates policies, institutional and legal arrangements, technologies, and data that enable sharing and effective usage of geographic information [4]. The degree of SDI development strongly correlates with the development of the information society in general, use of information technology by the population, and the diffusion of the internet [5]. Figure1 illustrates the conceptual SDI model.

This research is underlying the core components of the Egyptian Standardized Spatial Data Infrastructure. ESDI consists of more than a single dataset or database; it hosts different spatial data collections and its related attributes, it provides sufficient metadata and specifications, as well as a means to migrate, quality control, and data access capabilities.

Address for correspondence:

amrhali@feng.bu.edu.eg

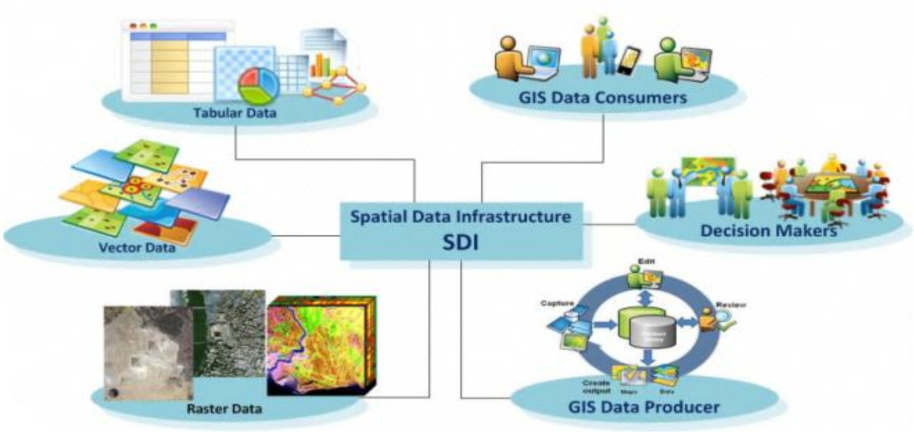


Figure1. *Conceptual SDI Model*

RESEARCH METHODOLOGY

The use of descriptive analytical method in this research allows first to define the SDI, its importance, and components, then we are going to focus on the core mechanisms of the establishment of the ESDI. Figure 2 summarizes this methodology.

Figure2. *Research Methodology*

(Data "different datasets, levels, themes, scales", QC, migration, accessibility, standard modelling)

Traditionally, the applied datahave different standards, documentation, and accessibility, which results in inconsistencies that may create major inefficiencies and limit effectiveness. It also increases the time, effort, and cost to migrate the heterogeneous datasets.

SDI represents the basis for strategic planning in numerous fields. It establishes linkage and balance between economic, environmental, and social capital in order to improve the basis for societal response [6]. The value of the spatial information and the effectiveness of the decision making are much correlated to the quality, completeness, reliability, update, and availability of the spatial information. Consequently, SDI considered as an infrastructure for the spatial data manipulation which is access, management, integration, analysis, and communication. So, standards form a key factor underpinning the management of these data.An efficient SDI is an important asset in societal decision and policy making, operational governance, citizen participation processes, and private sector opportunities [7].

The key capabilities of an SDI are (a) Enriching spatial data sharing to a wide range. (b) Permitting integration of geographically distributed spatial information. (c) Enabling collaboration by multilateral information exchange and synchronization. (d) Allowing autonomous organizations to develop interdependent relationships in a distributed environment. (e) Facilitating the definition and sharing of spatial semantics [8].

In general, the term SDI is often used to denote the relevant base collection of technologies, policies, and institutional arrangements that facilitate the availability of and access to spatial data [9]. Figure 3 identifies the SDI components and underlying principles of the proposed Egyptian SDI.

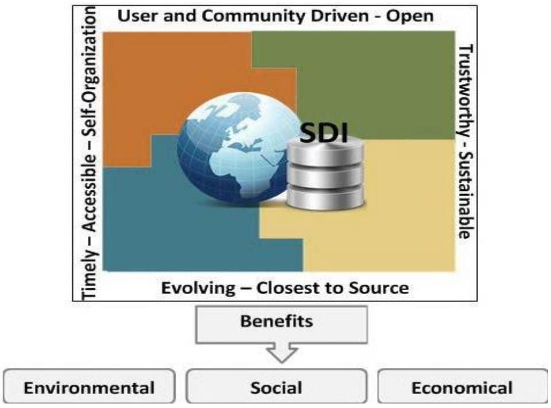


Figure3.*SDI Components and Guiding Principles*

Another aspect of utmost importance that affects the usage of spatial data is the interoperability. It includes three types of levels obstruct, which are (1) cross-border for matching different datasets, (2) cross-sector for combining datasets from different sources and different content, and (3) cross-type for combining different data storage formats as raster with vector data [10].

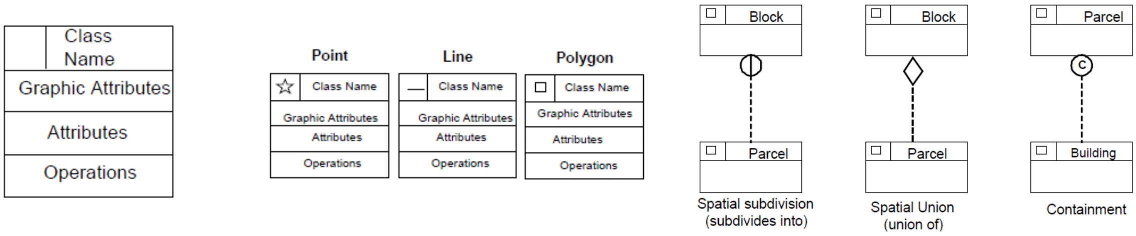
The primary components of SDI can be briefly described as follows (a) data framework that provides the model that hosts the set of continuous and fully integrated standard spatial data. Framework data are expected to be widely accessible, (b) Metadata which are technical documents that have been developed to address specific interoperability challenges. (c) Policies that includes the strategic and operational level that facilitates the development or use of an SDI. Strategic policies address high-level issues for organizations, while;operational policies address topics related to the lifecycle of spatial data and help facilitate access to and use of spatial information. (d) Technological architecture of an SDI which is composed of a network of physical servers that provide web services and data via these services, in such a way that an application can be developed tomake use of these services using the internet that represents the "highway" between data and services[11].

SDI Data Model

It is an abstract and partial representation of aspects of the world that can be manipulated to analyse the past, define the present, and to consider possibilities of the future [12, 13]. For many years database modelling and design have been studied, where two major types of data models used in many current DBMS are the relational data model and object oriented data models [14].

The relational database management systems are suitable for many applications that deal with simple structured data, while spatial database model require complex structure, as it integrates data from different resource into a consistent homogenous structure. Spatial data requires flexible data models that achieve multiple tasks which are (1) sophisticated abstraction of real world geometry, (2) representation of the same data at diverse conceptual levels of details, and different scale, (3) merging data from multiple data sources with different precision and accuracy, and (4) management of historical data.On the other hand, object oriented database modelling technology has impacted many fields for more than two decades. The Object based data model treats spatial data as objects that can represent a spatial feature as road, building etc.., and represent the coordinate system. Object oriented data models are defined to capture more semantics than relational models [15].

The object-based data model is different from the spatial data model in two important aspects. First, the object based data model stores both; spatial and attribute data of spatial feature in a single entity. Second, the object-based model allows a spatial feature (object) to be associated with a set of properties and methods. A property describes an attribute or characteristics of an object. A method performs a specific action [13].The object oriented data model is built on the four basic concepts of abstraction, which are classification, generalization, association, and an aggregation [15]. An additional significant concept in object oriented modelling is inheritance; that defines the relationship between a super class and a sub class [14]. Figure 4 defines the representation of the basic elements of object oriented model.



a. Class representation b. Geo-Object c. Relationship

Figure4. *Research Methodology Illustrationof Object Oriented*

The UML modelling of the spatial database differs from the well-known principle of modelling, where it includes some requirements in close association with the basic elements definition such as entities, relationship … etc. These requirements can be classified into some categories, which are (a) constraints that should be attached to objects (e.g., limiting the value of an attribute to a certain range, (b) associations which include topological relationships, that define the geometrical relationship (e.g., touch in, cross, overlap and disjoint), or the metric specifications (involving distance precision and the absolute position of feature in a reference coordinates system), semantic relationships, and relationships to indicate that an object is composed of other objects, where associations and constraints are related to the data quality, and (c) spatiotemporal properties that include the spatial requirements (reference coordinates system, geometry representation); data source timing and historical changes; attributes information, and a unique identifier that states as indexing, as well as spatial indexing [16].

There is no doubt that SDI system has to be equipped with appropriate index structures in order to generate efficient execution for queries comprising attribute and spatial data types [17]. An index provides pointers to the rows in a table that contains a given key value. A regular index stores a list of IDs for each key corresponding to the rows with that key value [18]. The implemented spatial indexes are conceptually built using B-trees, where the indexes must represent the 2-dimensional spatial data in a linear order of [19]. The index-creation process decomposes the space into a multiple-level grid hierarchy. These levels are referred to as level 1 (the top level), level 2, level 3 ….etc. Each successive level further decomposes the level above it, so each upper-level cell contains a complete grid at the next level. On a given level, all the grids have the same number of cells along both axes (for example, 4x4 or 8x8), and the cells are all one size as illustrated at Figure 5 [18].

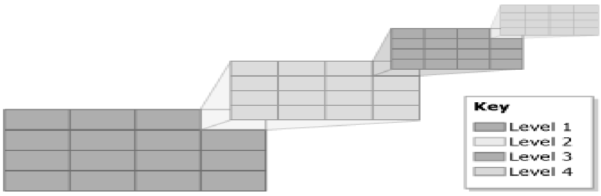


Figure5. *Illustration of decomposition for the upper-right cell at each level of the grid hierarchy into a 4x4 grid*

The algorithm of using B-tree to find the value of an object with key=k is represented in Figure 6.

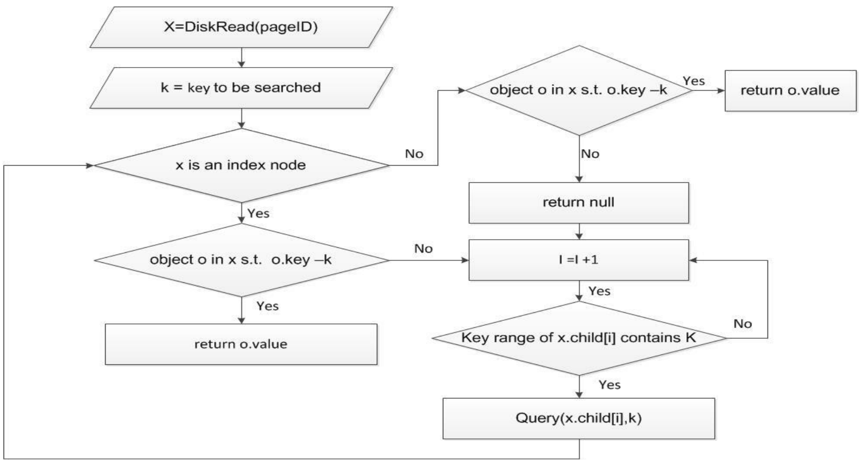


Figure6. *Query Algorithm Flowchart*

Another significant component is metadata, which provide a summary with detailed elements pertaining to a particular SDI. These elements are (a) identification info that provides basic information about the data sets including (title, geographic data coverage, information sources and currency of data), (b) data quality information which contains information about the quality of data set, including (positional and attribute accuracy, completeness, and consistency), (c) spatial data organization information which comprises information about the data representation in the data set, such as (method for data representation and number of spatial objects), (d) spatial reference information that contains description of the reference frame for and means ofencoding coordinates in the data set, such as (the parameters for coordinate system, horizontal and vertical datums), (e) feature catalogue which encompasses information about the content of the data set, such as(the entity types and their attributes and the standard codes of fields), (f) distribution information that incorporates information about sharing the dataset [20].

Spatial Data Migration

Due to the increasing amount of existing data and data flow from different sources, it is required to move the existing - and future - heterogeneous data to the new developed, unified, integrated, and centralized SDI with a universal schema. Spatial data migration is the process of transferring data between different storage types, or data formats, which accomplished through three separate steps. Mapping process is used to define the relationship between the data source and SDI. It involves combination of multiple name fields into one field, data mapping from one or multiple representations to single representation [21]. The next step is the transforming processes that adopt the data model with the target SDI. Data transformation includes different aspects that defined as(a) data clean-up that applied in order to remove errors within a dataset, (b) data merging/ conflation that involves bringing multiple datasets together into a common framework and combination of different sources of spatial data attributes, (c) data verification for quality assurance purposes, (d) data translation for the conversion of spatial data from one format directly to another with no intended change in structure or schema, and (e) spatial data transformations for coordinate systems and projections conversion. The last step of data migration is the load function which writes the resulting data to the target SDI [22]. By using these spatial data migration techniques, disparate spatial datasets maintained by multiple organizations can be combined into a common data model. Figure 7 illustrates the conceptual workflow of migration process.

Figure7. *Illustrates the migration process*

We should emphasize that migrating data into the system will not enhance or degrade the overall quality; in other words, the migrated data must be consistent in itself and consistent with data already in the SDI. Therefore, a data quality process will be added above the data migration process.

Spatial Data Quality

It is the degree of excellence exhibited by the data towards the actual scenario in-use. It is generally thought of as a multi-dimensional concept and is most commonly referred as "Fit-for-use" [23].

The data quality elements are (a) completeness which is a measure of the presence and absence of features, their attributes and relationships, (b) logical consistency that measures the degree of adherence to logical rules such as (data structure, attribution and relationships), (c) conceptual consistency for measuring the degree adherence to rules of the conceptual schema; it includes domain consistency that defines the adherence of values to the value constraints, as well as the format consistency degree to which data is stored in accordance with the physical structure of the data set, (d) physical consistency to define the topological correctness and geographic extent of the database, and (e) positional accuracy that measures how well each spatial object's position in the database matches reality [24, 25]. Table 1 shows the matrix of quality elements of spatial data function of the spatial data component [26].

Table1. *The Matrix of Quality Elements of Spatial Data*

	Space	Time	Attribute	Scale	Relationships
Accuracy	Positional Accuracy	Temporal Accuracy	Attribute Accuracy		
Consistency	Logical Consistency	Temporal Consistency	Codes Consistency		Topology
Completeness				Completeness	

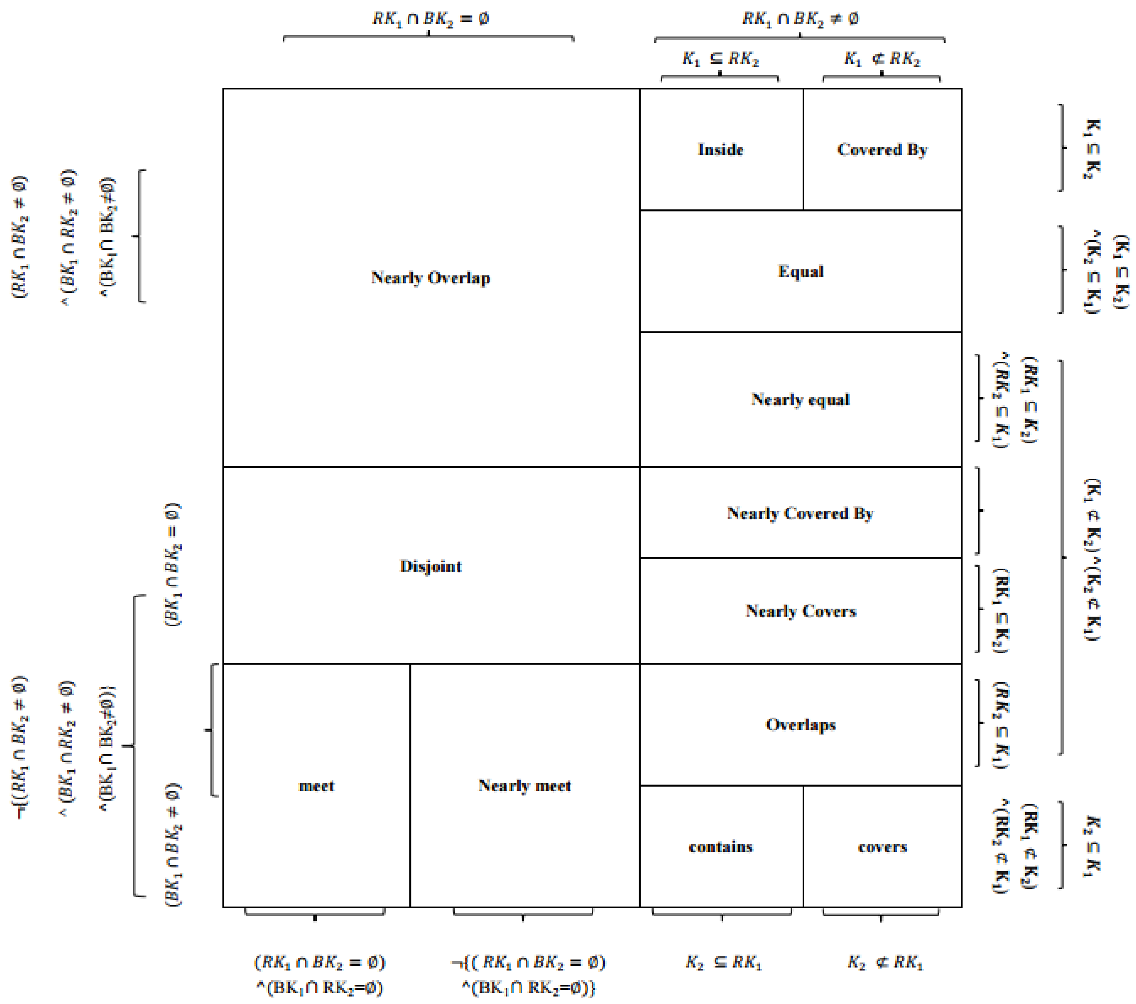


Figure8. *Conceptual Equations of Topological Relationships*

Asshown in Figure 8, topological relationship can be represented based on spatial object components, because any uncertain spatial objects K can consist of three components, that is the core RK, crust BK and exterior EK. Thus, the topological relationship conceptual equations between any two objects K1 and K2 can be represented as one of the thirteen basic relationship types [26].

One of the standard methods to measure the positional accuracy is root-mean-square error (RMSE). Horizontal RMSEr can be represented as shown at Equation (1) and the horizontal accuracy according to the National Standard for Spatial Data Accuracy (NSSDA) defined as Equation (2)

$$RMSE_r = \sqrt{\frac{\sum_{i=1}^{n}[(X_{data\,(i)} - X_{Check\,(i)})^2 + (Y_{data\,(i)} - Y_{Check\,(i)})^2]}{n}} \quad (1)$$

Where:

$X_{data\,(i)}, Y_{data\,(i)}$: the coordinates of the ith check point in the dataset.

$X_{Check\,(i)}, Y_{Check\,(i)}$: the coordinates of the i^th^ check point in the independent source of higher accuracy.

N: the number of check points tested.

$$\text{Accuracy}_r = 1.7308 \cdot \text{RMSE}_r \quad (2)$$

Similarly, vertical RMSE as well as the vertical accuracy according to NSSDA expressed in Equation (3)&(4) respectively.

$$RMSE_z = \sqrt{\frac{\sum_{i=1}^{n}(Z_{data\,(i)} - Z_{Check\,(i)})^2}{n}} \quad (3)$$

Where:

$Z_{data\,(i)}$: the vertical coordinates of the i^{th} check point in the dataset.

$Z_{Check\,(i)}$: the vertical coordinates of the i^{th} check point in the independent source of higher accuracy.

N: the number of points being checked.

$$\text{Accuracy}_Z = 1.9600 \cdot \text{RMSE}_Z \quad (4)$$

The relationship between the allowable accuracy in meter and the map scale according to the national standard represented at Table 2.

Table2. *Allowable Accuracy Standards for map scales*

Scale	Accuracy (M)	Scale	Accuracy (M)	Scale	Accuracy (M)	Scale	Accuracy (M)
1:100	0.025	1:500	0.125	1:2500	0.650	1:1000	2.550
1:250	0.065	1:1000	0.250	1:5000	1.270	1:25000	6..350

See [27].

Quality Control (QC) is a series of analytical measurements (tools) used to assess the quality of the data [28]. QC process can be summarized in three basic stages, (1) The review stage for identifying the errors; two key methods are available for finding errors which are automatically through automatic QC check or manually through a visual approach, where checking process focuses on ensuring that the collection of components (perhaps comparable to the SDI Model) meets the specified requirements. (2) Correcting the errors, where the concerning personnel must correct the raised errors as per the generated report that highlights the details of such errors. (3) Verification to validate that the errors are corrected and to ensure that no new errors were introduced in the process of fixing errors [29].

The two main approaches of QC are automatic and visual. The automated QC defines all major standards violations up front, because these review processes are automated with little human interaction, this is the fastest form of quality control. This means it is often possible to validate 100% of data. The automated QC includes two stages that described as follows:

Stage 1: Initial QC Checks that includes (a) verifying spatial reference to check that the delivered data have the definition of (the projection, coordinate system, datum, domain extent) as specified at SDI, (b) checking the database structure (schema) for ensuring that the delivered data follows the schema of SDI, (c) validating the field properties to examine the data type for each field of the feature classes; the subtype’s definition; the relationships between feature classes; and the assignment of constraints and ranges of attributes are matching the SDI Schema, and (d) confirming the coded values that inspects the constraints, ranges, and codes values with respect to the SDI schema.

Stage 2: Checklist Verification which contains (a) attribute checking for ensuring that the fields which don’t allow null are not missing its values; as well as the completeness of the fields as stated below at the acceptance criteria; moreover, checking fields for unique values as coding fields as defined at the data dictionary of the SDI, (b) check the geometry for badly formed geometry such as (duplicate features, short segments, null geometry, self-intersection, unclosed rings, empty parts, and (c) topology validation: to ensure that the defined topological rules are 100% valid [30].

The second type is the visual QC that takes advantage of what the human eye does well; find missing features or things that are out of place[25]. Visually inspecting data can detect systematic errors such as anoverall shift in the data caused by an unusually high RMS value as well as random errors such as misspelled text. Visual inspections can detect the presence of an erroneous date, the absence of data, or positional accuracy of data. Visual QA can be performed using on-screen views for checking for missing features, misplaced features, and registration errors. Many techniques can be applied for visual inspection as symbols and labelling [30]. The following diagram Figure 9 illustrates the quality control workflow.

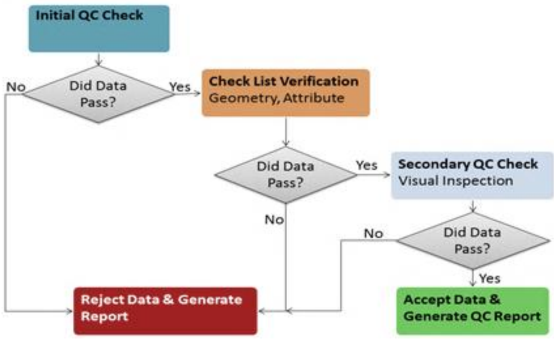


Figure9. *Quality Control Workflow*

The probability of errors is given by the Equation (5)

$$\emptyset = P(Y = 1) = \frac{M}{N} \tag{5}$$

Where, N is the total objects numbers, M shows at least one error which is unknown and subject to be estimated by evaluating the sampling results.

With a specific value y_i of the variable Y for the i^{th} investigated objects, the number of errors m in a sample of size n easily can be calculated by summation as in Equation (6)

$$m = \sum_{i=1}^{n} y_i \tag{6}$$

As shown at Equation (7), the probability of drawing exactly m erroneous objects in a sample of size n is given by the hypergeometric distribution.

$$P(X = m) = \frac{\binom{M}{m} \cdot \binom{N-M}{n-m}}{\binom{N}{n}} \tag{7}$$

$$with\ 0 \leq m \leq n \quad m \leq M\ and\ n - m \leq 0$$

Where, the random variable X holds for the number of errors in a given sample that depends on the parameters N, n and M. For numerical evaluation, Equation (8) should be used.

$$\prod_{i=0}^{m-1} \frac{n-i}{m-i} \cdot \prod_{i=0}^{m-1} \frac{M-i}{N-n+m-i} \cdot \prod_{i=0}^{n-m-1} \frac{N-M-i}{N-i} \tag{8}$$

See [26].

Egyptian Standard Spatial Data Infrastructure Design

In this research, the proposed designed ESDI focuses on the core components of the SDI, which includes the SDI database model, Metadata, Spatial Data migration as a tool for interoperability, as well as, quality specifications and automated quality control utility. The study area (Southern Upper

Egypt Region) is located at the Southern of Egypt and composed of Sohag, Qena, Luxor, Aswan, and Red Sea Governorates. The geographical extents of the study area of Southern Upper Egypt Region are surrounded by thick border as shown in Figure 10.

Figure10. *Governorates of Southern Upper Egypt Region (Study Area)*

ESDI Model Design

The developed SDI has common principles, concepts, and methods to establish better integrity of spatial information. It targets cross-cutting issues, such as structured data presentation to guarantee data sharing, reliability, and integrity. It also guarantees flexibility to enable information interchange, and linking from multiple sources to a well-defined reference.

The three applied approaches that lead to a multiplication of spatial data are listed as the following:

- Multi-scale representation where the entities of the real world is described with varying levels of details, where the process of generalization involves reducing the amount of details in the representation of information.
- Multi-thematic representation that depends on the perspective, the same spatial data is represented in different forms. Each community emphasizes those properties of the data that are interest to a specific field and each representation outlines a specific thematic; where the geometrical features are the same but the information derived from this data are different for each scenario.
- Multi-temporal representation where the world changes overtime. Multiplicity of information that relates to the same place in different moments of time offers enormous potential for gaining a better understanding of our world [1].

The proposed architecture serves these multiplications through object oriented modelling schema. Figure 11 shows the designed schema of the proposed SDI model while, Figure 12 shows an abstract of the designed SDI model.

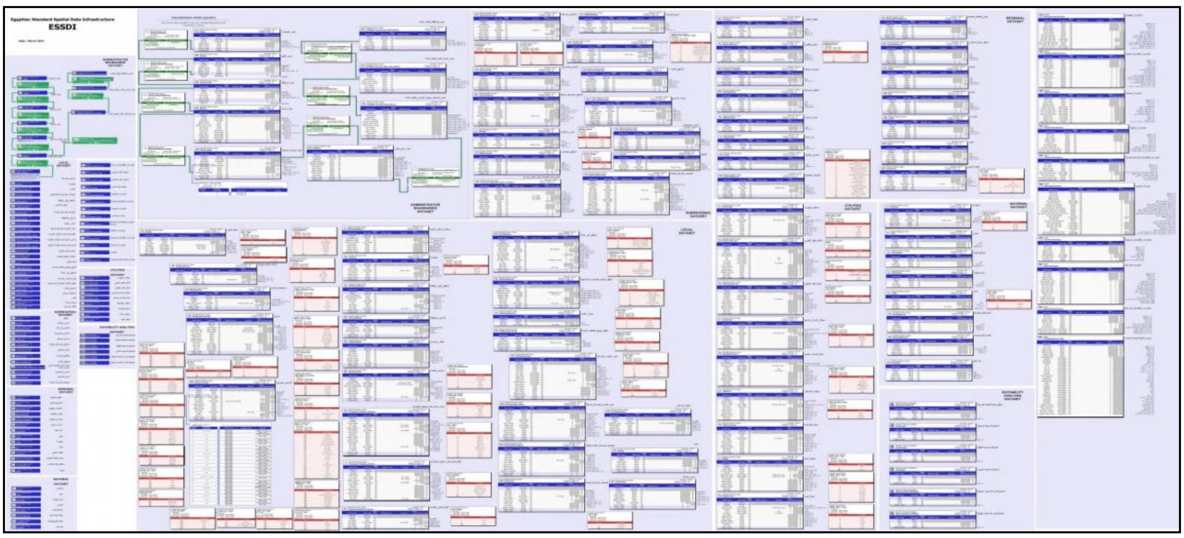
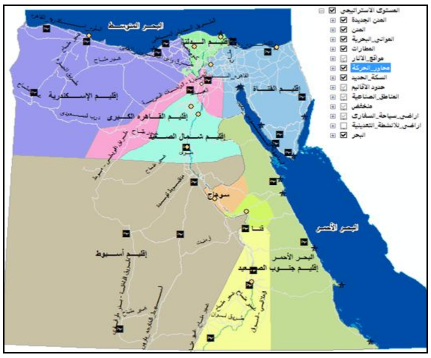


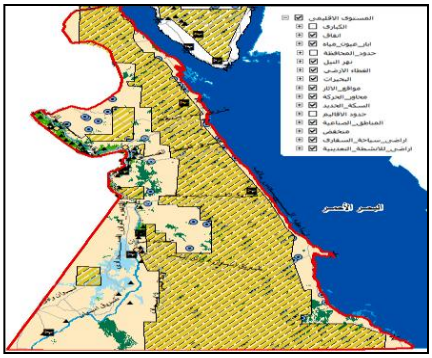
Figure11.*Schema design of the SDI Model*

Figure12. *Abstracted SDI Model*

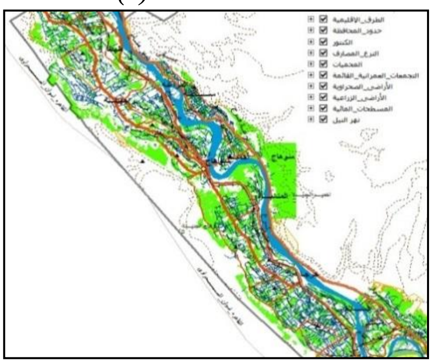
ESDI applies multi-scale approach through defining the different levels of data relevant to the Egyptian administrative level starting from the national level up to the regional level, governorate level, district level, city level, and finally the village level with the ability of cross cutting between levels. Figure 13 represents this approach.



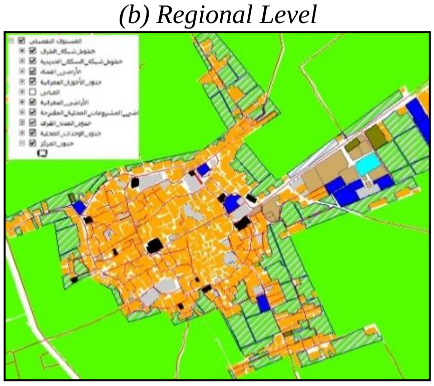
(a) National Level



(b) Regional Level



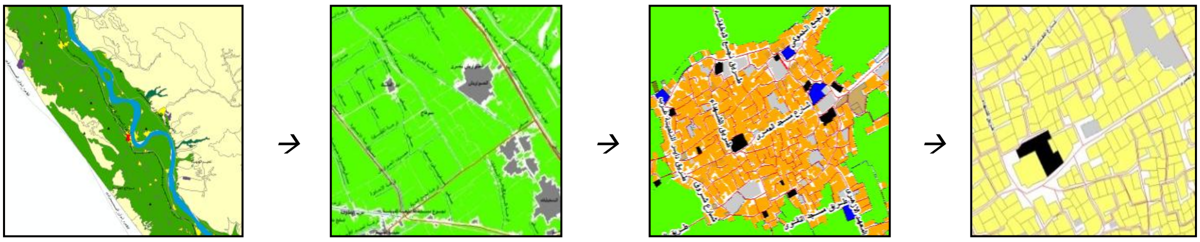
(c) Governorates Level



(d) Local Level (District/City/Village)

Figure13. *Multi-level representation*

Likewise, the approach that associates different levels of detailed is called a multi-scale representation, where the generalization concept is applied. In the case of describing an urban coverage settlement, as seen in Figure 14, a detailed description includes single buildings and all street networks in the area, a less detailed one provide only blocks of buildings and main roads, while in small scale all blocks are representing as one built-up area. The less detailed representations will include only a point that represents the location of the settlement and the major road network.



(a) 1:400,000 *(b) 1:100,000* *(c) 1:10,000* *(d) 1:2,500*

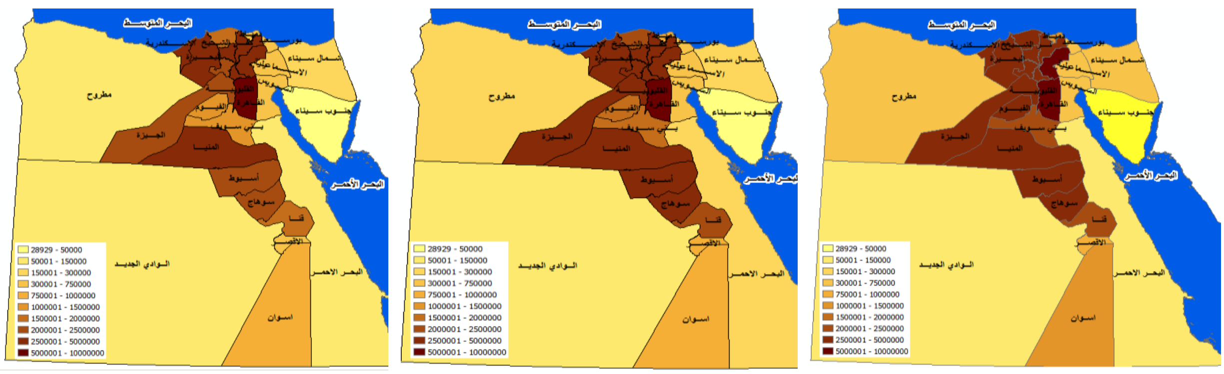
Figure14. *Multi-scale representation*

While, the multi-thematic representation represents the same geometry with dissimilar description, each description is valid but the information derived from it is different for each scenario. For example, the building geometry can be regarded as a land use, map, building status map, or building heights map as seen at Figure 15. Each map outlines a specific thematic field in other words classification of spatial object.

(a) Land use Map *(b) Building Status Map* *(c) Building Heights Map*

Figure15. *Multi-thematic representation*

Multi-temporal representation is a multiplicity principle that links a spatial object in a specific moment of time with its predecessor and successor. For example, Figure 16 represents the population map from year 1986 till year 2006.



(a) Population Map : 1986 (b) Population Map : 1996 (c) Population Map : 2006

Figure16. *Multi-temporal representation*

In addition to the data model, it is needed to document each data class. This is called a data specification /metadata that contains the data model description and other relevant provisions concerning the data, such as coding values, data quality requirements, metadata for data consistency and reusability [32]. **Figure17.**Figure 17 is a sample of metadata for class description of the ESDI Model.

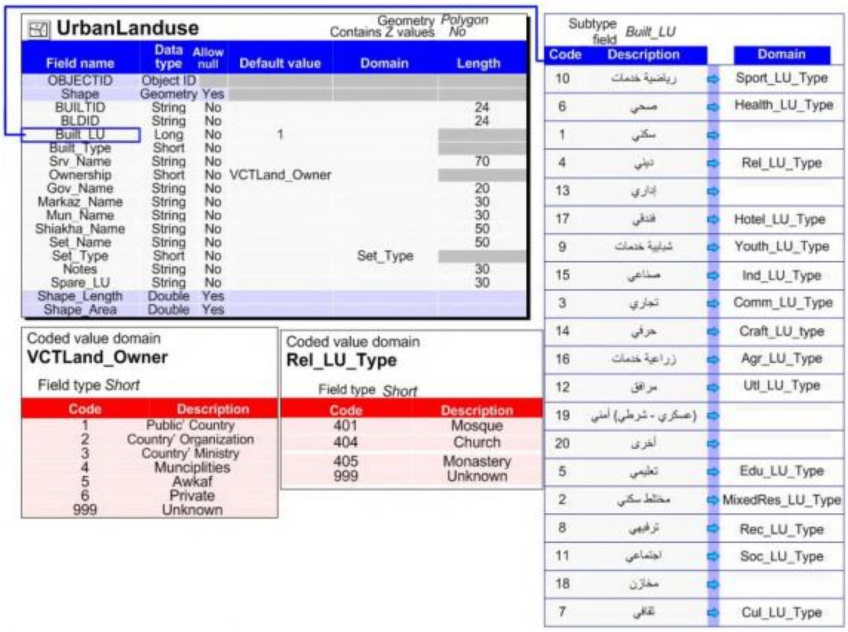


Figure17.*Sample metadata of Land use class*

ESDI Spatial Data Migration

The migration process means the ability to integrate spatial data from disparate sources to the standard data model. One way of achieving such migration would be is to comply with the dataset, so each dataset is selected as a target model and all other data is transformed to comply with its metadata [33].

In our case study, the villages and cities databases of Southern Upper Egypt region's governorates, and the environmental databases for the regions are migrated. The assessment of the existing data and the evaluation of the expected needs revealed that enumerating all possible formats of existing and intuitively future data sources is not reasonable. Therefore, building an ad-hoc data migration utility appeared to be the most feasible solution. This means that migration mapping can be defined during not only the design and implementation of the ESDI but also after the deploying. The mapping criteria between any data source to the ESDI will be applied through the developed utility. Consequently, the migration process will feature the advantages of being automated and will cope with the diversity of the data sources.

The idea is illustrated in Figure 18. We can see that, when we need to migrate a piece of data from any supported format, the loading unit will execute the job automatically as long as there is a mapping from the source format to the ESDI schema. If the source format is new and the system has not learned it, then the mapping unit will be used to define how the mapping should be done.

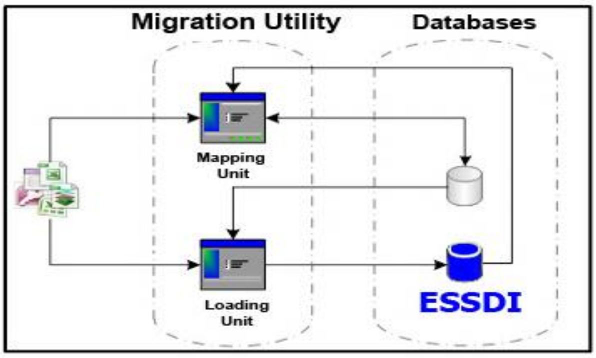
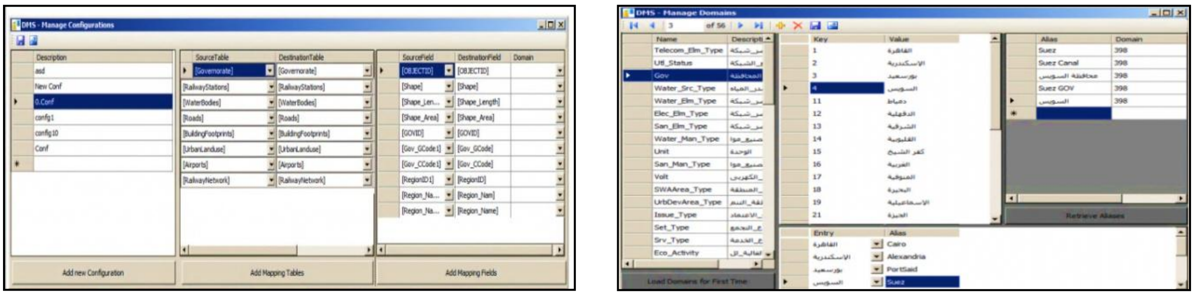


Figure18.*Illustrates the migration scenario*

Within the mapping unit we can define the table relation between source and destination schema. A source table can be related to one or more destination table and vice versa. Then, we can relate the fields of each table, similar to the table mapping, and in case of a coded field the related code description will be associated with it. Figure 19 represents the interface of tables / fields mapping as well as the codes description.



(a) Table and field mapping interface *(b) Codes description interface*

Figure19. *Mapping Unit Interface*

The last step of the migration process is the loading stage. Figure 20 shows the loading stage interface, where the application supports the standard applied formats (i.e. shapefile, geodatabase, Access, SQL, Oracle, Excel sheet, CSV file, etc).

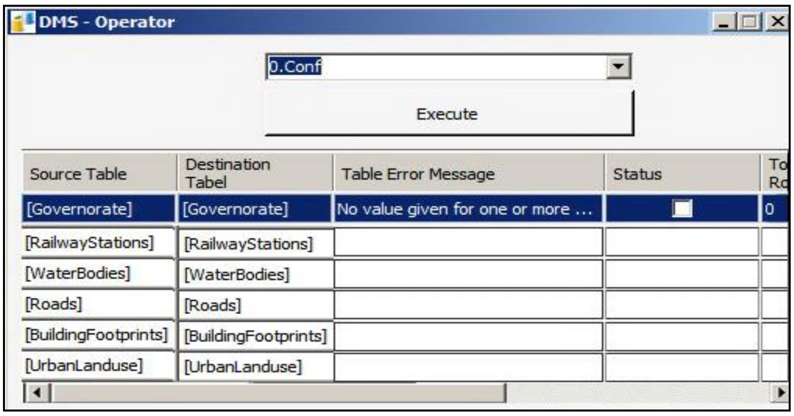
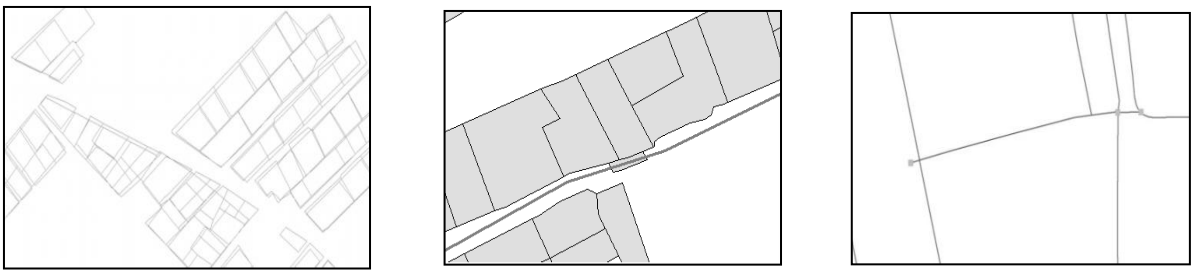


Figure20. *Loading Unit Interface*

Spatial Data Quality Design

Spatial data quality is a main concern where data is the core/backbone on which analysis and decision making relies. If data quality is not well addressed in terms of accuracy, completeness, consistency, and reliability, the analysis and calculations is expected to be irrelevant [34]. Quality is often regarded as a part of themetadata, however, in the case combine multiple sources, the quality of spatial information can naturally be assumed to play a more important role [35].

On this spot, the quality of a sample for Sohag governorate of villages, cities and governorate databases within the scale range between 1:2,500 and 1:100,000 have been examined. Figure 21 highlights the major discovered errors.



(a) Duplicate Buildings *(b) Road Network intersect Building* *(c) Road must not have Dangles*

(d) water bodies centerline isn't matching with the source layer

(e) Gaps between Land coverage features

(f) overlap between Buildings and Land coverage

Figure21. *Sample of the exposed errors*

The results emphasize the importance of the quality control procedures that should serve as a guideline/reference during data processing. So, the methodology that describe in comprehensive detail the necessary Quality Assurance (QA), Quality Control (QC), and other technical activities has been implemented to ensure the data quality. Furthermore, the automation of the quality controls process has been implemented through the development of a customized application. The sample of the generated reports for the three stages is represented as follows:

Stage (1) is seeking for the initial QC checks, which show the metadata generated report and highlights the missing items of the metadata (see Table 3). While Table 4 defines the feature layers which are not compatible with the final schema.

Table3. *Metadata Report*

Item	Status	Item	Status
Projection and coordinates system	Missing	Data producer	Missing
Data origin and Source	Missing	Methodology of data collection	Missing

Table4. *Compatibility with the Schema Report*

Regional Dataset	Environmental Dataset	Local Dataset	Utilities Dataset
• Regional Roads	• Agriculture Lands	• Railways Lines	• Water
• Contours	• Desert Lands	• Roads	• Electricity
• Natural Reserves	• Reclamation Lands	• Proposed Projects	• Sanitation
• Underground Water	• Military Areas	• Building Footprints	• Natural gas Lines
• Valley	• Fish Farms	• Vacant Lands	• Water Lines
• Sea	• Land Ownership	• Water Bodies	

Stage (2) is focusing on checklist verification, which states the percentage of the missing attributes within the feature layers (Table 5). A summary of the error percentages of the topological rules verification is shown at tables (Table 6)

Table5.*Missing Attribute Report*

Layer Name	Missing (%)	Layer Name	Missing (%)	Layer Name	Missing (%)
Planned Areas	2%	Road	2%	Water Wells	12%
Previous Urban Boundary	2%	Road Network	2%	Agriculture Lands	1%
Urban Boundary	5%	Railways Lines	3%	Desert Lands	1%
Land use	5%	Railway Network	1%	Existing Urban	2%
Vacant Lands	2%	Bridges	0%	Fish Farms	6%
Building	12%	Pollution Sources	4%	Industrial Areas	7%
Proposed Projects	10%	Underground Water	3%	Land ownership	20%
Water Bodies	1%	Valley	1%	Telecom Lines	17%
Regional Roads	2%	Water Floods	10%	Water Stations	14%

Table6. *Topological Rules Report*

Layer Name	Topological Rules	Error (%)	Layer Name	Topological Rules	Error (%)
Building Footprint	Must Not Overlap	2%	Electric Lines	Must Be Large Than Cluster Tolerance (0.1)	1%
Land ownership	Must be larger than cluster tolerance (0.1)	1%		Must Not Overlap	13%
Vacant Lands	Must be larger than cluster tolerance (0.1)	0%		Must Not Self Intersect	0%
	Must Not Overlap	14%		Must be single Part	6%
Proposed Roads	Must Not Have Dangle	0%	Sewer Lines	Must Be Large Than Cluster Tolerance (0.1)	0%
	Must Not Overlap	14%		Must Not Overlap	1%
	Must Not Self Overlap	1%		Must Not Self Intersect	0%
Unsafe Area	Must Not Overlap	0%		Must be single Part	0%
Urban Boundary	Must Not Overlap	2%	Telecom Lines	Must Be Large Than Cluster Tolerance (0.1)	1%
Waterbodies	Must Not Overlap	3%	Water Stations	Must Not Overlap	11%
Road Network	Must Not Overlap	0%		Must Not Self Overlap	2%
Agriculture Lands	Must not overlap	0%		Must Not Self Intersect	0%
Agriculture Lands	Must not overlap with Buildings Footprint	43%		Must be single Part	2%
Waterbodies	Must not overlap with Urban Land Use	51%	Water Stations	Must Not Overlap	2%
	Must Not Overlap With Vacant Land	23%		Must Not Self Intersect	0%
	Must Be larger than Cluster Tolerance (0.1)	0%		Must be single Part	8%

At stage (3), the visual QC took place, where Table 7 represents a part of the generated report which is related to the positional accuracy of the map of scale 1:5000 through visual inspection and comparing with satellite image as ground truth.

Table7. *Positional Accuracy*

P_ID	Δx	Δy	Δx^2	Δy^2	$\Delta x^2+\Delta y^2$	P_ID	Δx	Δy	Δx^2	Δy^2	$\Delta x^2+\Delta y^2$
1	-0.1	-0.13	0.01	0.02	0.03	6	0	-0.23	0	0.05	0.05
2	0.2	-0.15	0.04	0.02	0.06	7	-0.1	-0.24	0.01	0.06	0.07
3	-0.2	-0.16	0.04	0.03	0.07	8	-0.1	-0.27	0.01	0.07	0.08
4	-0.2	-0.18	0.04	0.03	0.07	9	-0.1	-0.28	0.01	0.08	0.09
5	-0.1	-0.21	0.01	0.04	0.05	10	0	-0.31	0	0.10	0.10
						----	----	----	----	----	----

The reported accuracy is equal to 0.37 m, which is within the range of map scale 1:5000 (See Table 2) according to the national standard, where this scale is appropriate for Government-to-Government (G2G) applicationsthat link government bodies to exchange services and documents.

CONCLUSIONS

- This research has tackled the development of Standardizing Spatial Data Infrastructures data (SDI) and its importance in Egypt as a key resource in the development of the Nation. There is a lot of economic potential that is locked away in spatial data holdings and this potential is realized by making the data widely available through a SDI. Additionally, SDIs are contributing to sound decision making, enhanced e-government applications, and location-based services.
- ESDI is of a vital importance for Governorates-to-Governorates (G2G) protocols that should be implemented to improve Decision Support Systems (DSS) methodology.

- The ESDI is a collection of spatial data framework through the proposed object oriented enterprise data warehouse models that is characterized with standardization. This standard provides sufficient documentation and metadata for the defined data model. Additional components are the means of exchanging data in harmonized and consistent matter through the ad-hoc spatial data migration utility along with the guarantee of the quality of data using the developed spatial data quality tool.
- The suggested enterprise database warehouse model represents the core component of ESDI. The model is implemented using object oriented concept. It is argued that such methodologies offer clear advantages over traditional methods such as E-R modelling, where object-oriented modelling allows incorporation of complex spatial data. Furthermore, the designed model addressed multiple levels of details and scales starting from micro to macro level, multiple themes for different vision of data representation, and preserving the data history through the multi-temporal representation. This model distinguished by the features of data indexing and spatial indexing along with data standardization through a review data specifications documents.
- As Data coming from multiple environments describing the real world from multiple points of view at multiple levels of details has to be presented. In spite of the complexity of the integration process, the migrated data must be coherent. While, ESDI sets up the framework of spatial data infrastructure integration to achieve the interoperability through the development of an ad-hoc migration application that follows the ESDI Model.
- Within our proposed ESDI schema, the data size of the migrated southern Upper Egypt has been reduced by 32%, where redundancy/conflict is discovered and eliminated by about 25%
- Meanwhile, a number of automatic quality procedures have been identified and developed, where these procedures are embedded into the data model.
- The data quality of Sohag Governorate as a case study matched the map scale of 1: 5000 which is appropriate for the G2G major applications.
- The importance of continuing the development of ESDI through the completion of the remaining components of strategic policies that enforce agreement with certain standards and procedures, and the technological architecture of software and hardware, as well as, joining the Africa National SDI of Africa through the government coordination.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to the General Organization of Physical Planning (GOPP), Cairo, Egypt for providingthe data applied at the case study.

REFERENCES

[1] KatalinTóth, Clemens Portele, Andreas Illert, Michael Lutz, and Vanda Nunes de Lima (2012) A Conceptual Model for Developing Interoperability Specifications in Spatial Data Infrastructure. Publications Office of the European Union.

[2] Richard Groot and John McLaughlin (2000) Geospatial Data Infrastructure, concepts, cases and Good Practice. Published by Oxford University, Press Inc., New York.

[3] HamedOlfat, Abbas Rajabifard, Mohsen Kalantari, and Chris Pett (2013) Design and Development of a Spatial Metadata Automation Framework Applied in Austral. Global Geospatial Conference, UNECA Conference Center, Addis Ababa, Ethiopia, November 4-8, 2013.

[4] Massimo Craglia and Andrew Johnston (2004) Assessing the Impacts of Spatial Data Infrastructures: Methods and Gaps. 7th AGILE Conference on Geographic Information Science, 29 April-1May 2004, Heraklion, Greece Parallel Session 1.1- "Spatial Data Infrastructure I" .

[5] Katalin. Tóth (2007) Data consistency and multiple-representation in the European Spatial Data Infrastructure. 10th ICA Workshop on Generalization and Multiple Representation – 2-3 of August 2007. Moscow.

[6] Amr Ali, EmanElAmir, Osman Hegazy, and Mohamed Noureldien (2012) Integrating Web Services with Geospatial Data Mining: Disaster Management for Road Accidents. Geoinformatica – An International Journal (GIIJ), Volume (2): Issue (1).

[7] Nama Raj Budhathoki and ZoricaNedovic (2007) Expanding the Spatial Data Infrastructure Knowledge Base. Research and Theory in Advancing Spatial Data Infrastructure Concepts (ed. Harlan Onsrud; Redlands, CA: ESRI Press, 2007).

[8] Prestige Makanga and Dr. Julian Smit (2010) A Review of the Status of Spatial Data Infrastructure Implementation in Africa. SACJ, No. 45., July 2010

[9] Douglas D. Nebert (2009) Developing Spatial Data Infrastructures: the SDI Cookbook. Retrieved from Global Spatial Data Infrastructure Association.

[10] Henri J. G. L. Aalders and Harold Moellering (2001) Spatial Data Infrastructure. Proceedings of the 20th International Cartographic Conference, Beijing, Volume 4, 2001, pp. 2234-2243

[11] Rio de Janeiro (2013) Spatial Data Infrastructure (SDI) Manual for the Americas. Tenth United Nations Regional Cartographic Conference for the Americas New York, 19-23, August 2013.

[12] Chinonye Cletus Onah, (2009) Spatial Data Infrastructure Model for Developing Countries: A case study of Nigeria.

[13] Keith C. Clarke, Bradley O. Parks, Michael P. Crane, (2002) Geographic information systems and environmental modelling. Prentice-Hall of India, New Delhi.

[14] Ramez Elmasri, Shamkant Navathe (1999) Fundamentals of Database Systems. Published by Addison Wesley Longman.

[15] Max J. Egenhofer and Andrew U. Frank (1992) Object Oriented Modelling of GIS. URISA Journal 4(2): 3-19.

[16] Jugurta Lisboa - Filho, Gustavo Breder Sampaio, Filipe Ribeiro Nalon and Karla A. de V. Borges, (2010) A UML Profile for Conceptual Modelling in GIS Domain, the CAISE'10 - 22nd International Conference on Advanced Information Systems Engineering.

[17] Hans-Peter Kriegel, Martin Pfeifle, Marco Pötke, Thomas Seidl, and JostEnderle (2004) Object-Relational Spatial Indexing, Spatial Databases Technologies. Techniques and Trends. Idea Group Inc., pages 49-80

[18] Rajeev Parida (2006) Principles and implementation of Data Warehousing. Published by Firewall media, New Delhi.

[19] Beng Chin Ooi, Ron Sacks-Davis, and Jiawei Han,1993: "Indexing in Spatial Databases"

[20] Drew Decker (2001) GIS Data Sources. Copyright © by John Wiley & Son.

[21] Manish Jain (2004) Data Warehouse, published by Manish Jain for BPB publications, New Delhi, India.

[22] Carmen Răduţ, (2013) The Quality of Data and Metadata in a data warehouse. RePEc:brc:journl:v:19:y:2013:i:1:p:36-40

[23] Shirlee Knight and Janice Burn (2005) Developing a framework for assessing information quality on the World Wide Web. Informing Science Journal Volume 8, 2005.

[24] Gray J. Hunter, Simon Jones, Arnold Bregt and Ewan Masters (2002) Spatial Data Quality. Advanced geographic information systems –vol. II.

[25] Antti Jakobsson, (2002) Data Quality and Quality Management - Examples of Quality Evaluation Procedures and Quality Management in European- National Mapping Agencies. Spatial Data Quality by Taylor & Francis, London

[26] Wenzhong Shi (2010) Principles of Modelling Uncertainties in Spatial Data and Spatial Analyses, Boca Raton: CRC Press/Taylor & Francis.

[27] V. Lynn Packard (2008) FGDC Federal Geographic Data Committee: National Geospatial Data Clearinghouse. Gratis http: //clearinghouse1.fgdc.gov/', Reference Reviews, vol. 22, no. 6, pp. 50-50.

[28] Suzie Larrivée, Yvan Bédard, Marc Gervais and Tania Roy (2011) New Horizons for Spatial Data Quality Research. 7th International Symposium on Spatial Data Quality.

[29] Michael S. O'Donnell, Timothy J. Assal, Patrick J. Anderson, and Zachary H. Bowen (2014) Geospatial Considerations for a Multi-organizational, Landscape-Scale Program. Journal of Map and Geography Libraries.

[30] Keith Mann (2006) QA/QC for GIS Data: Initial Quality Control Checks. Copyright © ESRI

[31] Keith Mann (2007) QA/QC for GIS Data: Visual In specification and Quality Control. Copyright © ESRI.

[32] EPA (2003) Guidance for Geospatial Data Quality Assurance Project Plans. One of the U.S. EPA Quality System Series.

[33] Pennsylvania Mapping and Geographic Information Consortium (2007) Pennsylvania Geospatial Data Sharing Standards (PGDSS): Spatial Data Migration. Pennsylvania Mapping and Geographic Information Consortium.

[34] Václav Talhofer, Alois Hofmann, Šárka Hošková – Mayerová, and PetrKubíček (2011) Spatial analyses and spatial data quality. AGILE 2011, April 18-22.

[35] Dr.Antti Jakobsson and Prof. Dr.LysandrosTsoulos (2007) The Role of Quality in Spatial Data Infrastructures. 23rd International Cartographic Conference, Moscow.

AUTHOR'S BIOGRAPHY

Amr H. Ali was bornin Cairo, Egypt.

He received the B.Sc., M.Sc. degrees in Surveying fromZagazig University-Egypt in 1983 and 1990, respectively. He holds his Ph.D. degree in Geodesy from the Technical University Graz, Graz, Austria in 1996. For over 30 years, Amr Ali has offered his deep knowledge and vast experience as an Associate Professor in Surveying Sciences at the Faculty of Engineering -Shoubra, Benha University. He has extensive experience in teaching Surveying, Geodesy Geodetic Sciences, Digital Mapping and Geographic Information Systems (GIS).He has numerous researches in Geodesy, navigation systems and Geomatics. Amr has also managed multiple academic projects in national and international levels. He has numerous researches in navigation systems, satellite geodesy, data warehouse, data mining, remote sensing and Geomatics.