IMPROVING THE CONSISTENCY BETWEEN TILES IN TILES BASED GIS - CASE STUDY: SURVEY DEPARTMENT 10K DIGITAL TOPOGRAPHICAL DATABASE

D.T.N.Jayasumana
Assistant Superintendent of Surveys, GIS Branch, Surveyor General's Office,
P.O.Box 506, Colombo-05, Sri Lanka
E-mail : jayasumana08064@alumni.itc.nl

KEY WORDS : Inconsistencies, Attribute, Tables, Structure, Coverage

ABSTRACT

Survey department of Sri Lanka is presently creating 1 : 10,000 (10K) digital topographical database. Data are stored in Arc Info coverage format based on tiles. Each 1 : 50,000 (50K) tile is subdivided into 25 numbers of 10K tiles. There are twelve coverages in each tile depending on the availability of data and the availability of features in real world. These coverages represent point, line, polygon features or combinations of points, lines and polygons.

Creation of 10K data is done by several ways. Existing 10K hardcopy maps are scanned and digitized on screen in AutoCAD platform for conversion them into Arc Info coverages. Digital cartographic model data in DGN format and field survey data collected using GPS too are converted into Arc Info coverages. Data conversion process is presently carried out based on tiles since source data are available in tiles. Due to conversion of data is done based on tiles, some inconsistencies could be observed between tiles. It will make problems when data in several tiles are to be combined and analyzed. Those inconsistencies need to be identified and measures need to be taken to avoid or minimize the inconsistencies.

Existing 10K coverage data stored in tile base were studied to examine the possible types of inconsistencies between tiles. From this study, the structure of the available data, different field names and their size definitions associated with each coverage were identified. Further, how different attribute fields are organized in the attribute tables was examined. Then a method was developed to identify the availability of attribute fields, inconsistencies in field sizes and placement order of fields, automatically. Once the inconsistencies are identified, the method was then improved to correct those inconsistencies automatically as much as possible and to report the errors that can not be automatically corrected.

1. INTRODUCTION

Presently, the Survey department of Sri Lanka is creating 1 : 10,000 (10K) digital topographical database. Data are stored in Arc Info coverage format based on tiles. Each 1 : 50,000 (50K) tile is subdivided into 25 numbers of 10K tiles. Administrative boundaries, Buildings, Hydrographical features, Transportation network, Landuse, Place names, Grids, Reserves, Utility lines, Local authority boundaries, Mahaweli boundaries and Terrain are the twelve feature classes used to represent the real world in 10K dataset. Separate coverage is maintained for each feature class and the existence of these coverages in each tile depends on the availability of data and the availability of features in real world. These coverages represent point, line, polygon features or combinations of points, lines and polygons.

Different data sources are being used to create the 10K digital topographical data. Existing 10K hardcopy maps are scanned and digitized on screen in AutoCAD platform for conversion them into Arc Info coverages. Digital cartographic model data in DGN format are converted into Arc Info coverages. Field survey data collected using GPS too are converted into Arc Info coverages. AML programs run in Arc Info software are used to convert the CAD data into Arc Info coverages. Data conversion process is presently carried out based on tiles since source data are available in tiles. Due to conversion of data is done based on tiles, three types of inconsistencies could be observed between tiles. They can be categorized as structural, attribute and spatial inconsistencies. They will create problems when data in several tiles are to be combined and analyzed. Inconsistencies on structure of attribute tables of feature classes were identified in this study and the measures need to be taken to avoid or minimize the inconsistencies were proposed.

2. METHODOLOGY

2.1 Study on 10K database

When a data coverage is created, some of the fields in attribute table are generated automatically by Arc Info software and some attribute fields are added to describe the feature, depending on the feature characteristics. Arc Info software maintains two attribute tables separately, AAT for arcs and PAT for point or polygon features. When the attribute fields are added to a table, the name and size of the field should be similar in all the tiles, and they should be placed in a similar order in every tile. If the attribute tables of two tiles to be combined do not full fill those requirements, all the attribute fields may not remain in the combined output table.

List of standard attribute field names which also are known as items and placement order of them in attribute tables of selected Arc Info coverages in 10K database for point, arc and polygon feature classes are shown below.

PLACES PAT :

AREA,PERIMETER,PLACES#,PLACES-ID,GFCODE,SDCODE,NAME,YEAR,METHOD

ADMIN AAT :

FNODE#,TNODE#,LPOLY#,RPOLY#,LENGTH,ADMIN#,ADMIN-ID,GFCODE,SDCODE,YEAR,METHOD

ADMIN PAT :

AREA,PERIMETER,ADMIN#,ADMIN-ID,PROV,DISTRICT,DSAREA,GNDIVISION,YEAR,METHOD

TERRAIN PAT :

AREA,PERIMETER,TERRAIN#,TERRAIN-ID,GFCODE,SDCODE,ELEVATION,YEAR,METHOD

TRANS AAT :

FNODE#,TNODE#,LPOLY#,RPOLY#,LENGTH,TRANS#,TRANS-ID,GFCODE,SDCODE,TYPE,NAME,YEAR,METHOD

HYDRO PAT :

AREA,PERIMETER,HYDRO#,HYDRO-ID,GFCODE,SDCODE,NAME,YEAR,METHOD

Each item in attribute table has a *name, item width, output width* and an *item type*. *Item name* describes the item. *Item width* specifies the number of spaces (or bytes) used to store item values. *Output width* specifies the number of spaces used to display the item values. *Item type* is the data type of the item. For e.g. Character item is denoted by *C*, Integer item is denoted by *I* and decimal numbers are represented by *F* in item definition. For the item types which hold decimal numbers will have an additional definition of *No. of decimals.* Standard size definitions of selected attribute fields in 10K Arc Info attribute tables are given in Table 1.

Item Name	Item Width	Output Width	Item Type	No. of Decimals
GFCODE	5	5	C	
SDCODE	3	3	I	
NAME	50	50	C	
YEAR	4	4	B	
METHOD	2	2	B	
TYPE	4	4	C	
ELEVATION	4	8	F	2
PROV	2	2	C	
DISTRICT	12	12	C	
DSAREA	30	30	C	
GNDIVISION	30	30	C	

Table 1 : Standard size definitions of attribute fields in 10K database

2.2 Identification of inconsistencies and taking actions on inconsistencies found

For each feature class in 10K database, following procedure is adopted for detecting inconsistencies and correcting them and is shown in Figure 1. AML program was developed to process the procedure in Arc Info environment. First the item placement order in attribute table was read. Then if an item exists in feature attribute table, the size definition of the item (*item width, output width*, *item type*) was read. The placement order of the items and item size definitions of all the items in attribute table were then compared with standard placement order and standard size definitions. If item placement order does not match with standard order, it is no doubt that entire attribute table need to be arranged in standard order. Altering only the incorrect attribute size definitions is insufficient because it affects the item placement order in the table. When dealing with item size definitions in Arc Info software, it is not permitted altering the size of *item width*, but it allows the alternation of all other item size definition values. So when it is needed to alter the item size definition, the item needs to be re-added to the attribute table, which causes alternation in item placement order.

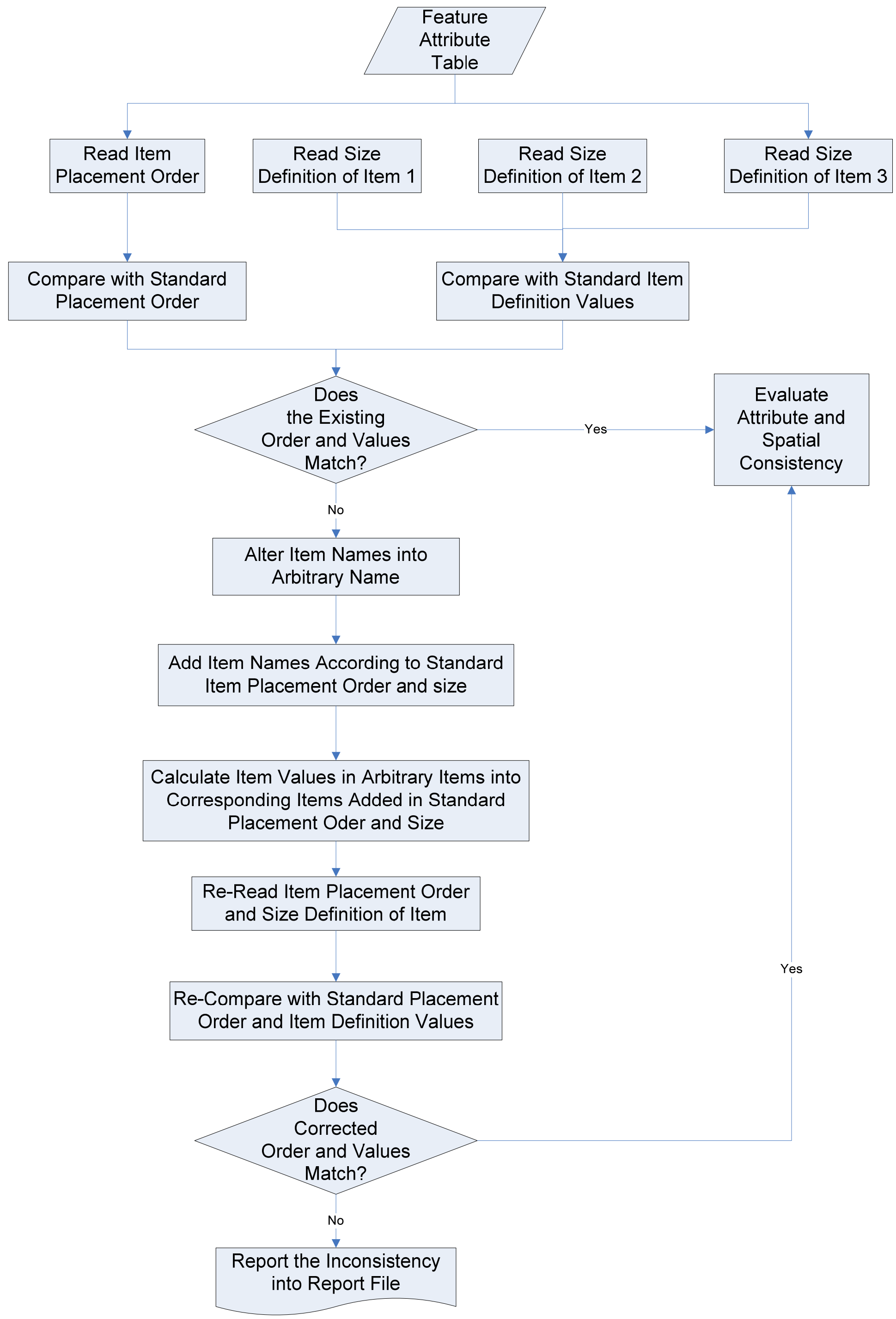

Figure 1: Procedure adopted for detecting and correcting inconsistencies

In the process of altering the item placement order and the size, care should be taken to avoid data lose. The data values available under the inconsistent item need to be assigned to the corrected item. So before dropping the inconsistent item, its data values need to be assigned to a temporary item which makes it possible to copy them back to the corrected item in the table. In this inconsistency detection and correction process, it is possible to detect the availability of any items other than the standard items that should be in the table. But, difficulty arose in separating the additional items from standard ones and therefore they could not be dropped automatically from the attribute table. As a solution for this matter the process was developed to report the inconsistencies that can not be automatically corrected and those inconsistencies need to be corrected interactively.

3. CONCLUSION AND RECOMENDATIONS

Availability of consistent data set will give many benefits to the users. This process of detection and correction of structural inconsistencies between tiles without losing the data compiled will make the first step of improving the consistency between tiles in a tiles based topographical database. Once it is possible to combine the data in adjoining tiles with their attribute values, attribute consistency and spatial consistency can be evaluated. Automation of identifying inconsistencies between tiles and correcting them automatically (at least for a certain extent), will helpful to improve the data quality in efficient manner.

Following recommendations can be proposed from this study on 10K topographical database.

- Use a standard set of AML scripts to convert CAD data into Arc Info coverages
- Avoid adding or dropping items in attribute tables manually
- This process of detection and correction of inconsistencies in structure can be utilized to ensure the structural consistency of the data compiled
- As further studies, a method need to be developed to evaluate the consistency of attribute values and to detect and correct spatial inconsistencies