# REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188



**1. AGENCY USE ONLY** *(Leave blank)*

**2. REPORT DATE**
25-29 March 1991

**3. REPORT TYPE AND DATES COVERED**
Scientific Paper

**4. TITLE AND SUBTITLE**

Spatial Data Transformation: Feature Attribute Conversion Issues & Practical Experience

**5. FUNDING NUMBERS**

**6. AUTHOR(S)**

MARK A. SITHER

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

U.S. Army Topographic Engineering Center
ATTN: CETEC-LO
Fort Belvoir, VA 22060-5546

**8. PERFORMING ORGANIZATION REPORT NUMBER**

R-185

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**



**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION/AVAILABILITY STATEMENT**

Approved for public release;
Distribution is unlimited.

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200 words)*

The need exists for continued development of spatial data transformation tools to provide increased capabilities for conversion of both format and content of locational coordinates, feature attributes, and ancillary information. Although individual geographic information systems (GIS) often provide a means of data import from and export to various other systems, such capabilities typically only permit transformation of the spatial data structure and format from one system to another. Existing feature attribute information is essentially transferred intact, subject to the constraints of the conversion software and other characteristics of each particular GIS; thus, reconciliation of differences due to variations in data description are left to the user. The tremendous growth in spatial data holdings constructed using a broad range of evolving systems, techniques and specifications has intensified the demand for enhanced data transformation capabilities to facilitate the integration of numerous existing data bases.

**14. SUBJECT TERMS**
1. Spatial Data Transformation
2. Feature Attribute Conversion
3. Spatial Data Exchange
4. Geographic Information System (GIS)

**15. NUMBER OF PAGES**
9

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| UNCLASSIFIED | UNCLASSIFIED | UNCLASSIFIED | |

SPATIAL DATA TRANSFORMATION:
FEATURE ATTRIBUTE CONVERSION ISSUES & PRACTICAL EXPERIENCE

Mark A. Sither
U.S. Army Engineer Topographic Laboratories
Telegraph & Leaf Roads, Cude Building
Fort Belvoir, VA 22060-5546
(202) 355-2858

## ABSTRACT

The need exists for continued development of spatial data transformation tools to provide increased capabilities for conversion of both format and content of locational coordinates, feature attributes, and ancillary information. Although individual geographic information systems (GIS) often provide a means of data import from and export to various other systems, such capabilities typically only permit transformation of the spatial data structure and format from one system to another. Existing feature attribute information is essentially transferred intact, subject to the constraints of the conversion software and other characteristics of each particular GIS; thus, reconciliation of differences due to variations in data description are left to the user. The tremendous growth in spatial data holdings constructed using a broad range of evolving systems, techniques and specifications has intensified the demand for enhanced data transformation capabilities to facilitate the integration of numerous existing data bases.

## INTRODUCTION

As the volume and diversity of digital geographic information continues to increase, the necessity for a transformation methodology to enhance the exchange of data between various geographic information systems (GIS) has grown in a proportional manner. The digital representation of a geographic feature includes both a locational geometric description as well as associated nonlocational information describing various characteristics or attributes of the feature. Efficient exchange of spatial data requires that certain "standards" exist in the treatment of locational description and the nonlocational or attribute information which describes the characteristics of various geographic features. Development and implementation of various standards currently underway; e.g., the Spatial Data Transfer Format (SDTF) in the United States and the Ordnance Survey National Transfer Format in the United Kingdom, will ultimately assist in the exchange of data among users. However, the evolutionary nature inherent within any new technology and the difficulties encountered while seeking to agree upon any universal data structure and attribute criteria will preclude the implementation of the majority of such standards for spatial data within the foreseeable future. Consequently, the need for tools to increase the transferability and utility of data among various systems will continue to grow.

## ISSUES IN FEATURE ATTRIBUTE CONVERSION

Fundamental differences exist between the management and processing of nonspatial data and that required for the handling of georeferenced spatial information. Thus, additional capabilities are necessary for geographic information processing and manipulation over and above those of basic storage and query typically found in traditional data base management systems and other similar data handling techniques. Interfaces have been developed for the conversion of the both structure and format of spatial information to allow the exchange of various types of data among the more established systems. Although primary concerns of spatial data exchange involve those of the structure, format and content of both the locational coordinate data and its related descriptive information, numerous other issues are confronted in the management and exchange of spatial data bases. Quality, accuracy, density, currency, consistency, completeness and historical lineage also should be considered. Global information, also of significance in the transfer of spatial data, includes the spatial coordinate reference system in which the data are stored, the geographic areal domain in which the data reside, and data dictionary in which various entities, associated attributes and attribute values potentially found within the data base are defined.

## DIGITAL TOPOGRAPHIC ENHANCEMENT PROGRAM (DTEP)

### Background

The problems associated with spatial information exchange are being addressed within many of the development efforts at the U.S. Army Engineer Topographic Laboratories (USAETL). Substantial prototype data has been created for use in developmental systems over the years, thus various defacto standards have evolved. As Digital Topographic Data (DTD) standards are established and such data becomes available, attempts to upgrade many of the prototype data sets for continued use are underway during the transition. Existing data can be used as a base source for the creation of data bases being produced under the new specifications. It also can provide an excellent foundation for construction of interim DTD upon reconciliation of variations in feature attribute classification encoding specifications.

The Digital Topographic Enhancement Program (DTEP) was initiated in May 1989. A portion of the effort included the development of interim capabilities for generation of DTD by Army terrain units in the field. The initial DTEP design plan prescribed that it would provide the capacity for spatial data creation from hard-copy Tactical Terrain Analysis Data Bases (TTADBs) and Planning Terrain Analysis Data Bases (PTADBs). Such design is similar to that required for the Digital Topographic Support System (DTSS), an Army field-deployable system. DTD would be created using a combination of techniques, including traditional manual digitization, use of automated scanning devices, and transformation of existing digital data into appropriate format and feature attribute encoding necessary for use with the Prototype DTSS (DTSS-P).

Primary spatial data creation for DTEP would draw from capabilities found within the TerraBase Military Terrain Analysis System originally developed at the U.S. Military Academy. The commercial GIS from Environmental Systems Research Institute (ESRI), pc ARC/INFO, would function as the principal geoprocessing component within the spatial data transformation process.

DTEP - Phase 1

The primary goal of the preliminary DTEP data transformation effort was to provide the capability for conversion of Surface Materials (Soils), Vegetation and Surface Configuration (Slope) information from the TerraBase raster format and internal feature classification system to that of the DTSS-P vector format and internal feature classification standard. ARC/INFO (DTEP-A/I), having been selected for DTEP because of its compatibility with the GIS scheduled for implementation in the fielded DTSS, provided a suitable, convenient apparatus to assist in the development of techniques for spatial data exchange. Such an approach also provided the capability for linkage of multiple iterations of feature attribute encoding to individual DTD coverages held within ARC/INFO's relational vector data base, thus facilitating the exchange of spatial information among various systems. Initial conceptual flow of data within the DTEP transformation effort is outlined in Figure 1.

Although a standard source of data had been selected for DTEP, the terrain feature information found on the 1:50,000-scale TTADB series and the 1:250,000-scale PTADB series, a lack of standard digital feature attribute encoding and storage specifications meant that numerous versions of DTD had evolved over the years. Consequently, determination of a proper set of feature attribute standards to be used within the project became a key element in the development of a variety of procedures for creation of new data and elimination of inconsistencies among existing digital data in order to assist exchange of information between systems operating with multiple data structures, formats and feature attribute encoding standards.

Various approaches to the synthesis of spatial information into systematized groupings of entities, attributes and associated values based upon similarities in data description have been set forth in the literature; e.g., "view integration" (Marble 1988) and "schema integration analysis" (Nyerges 1989). Such techniques employ both value judgement and mathematical procedures for the systematic examination of compatibility among multiple descriptions

**Figure 1**
**DTEP Data Transformation Effort - Phase 1**
**Data Flow Composite Overview**

```
1TerraBase ---------> |-----------> DTEP-A/I -> |-----> DTSS-P

2DIGIDATA-> DIGITAS-> [CAS]->        ARC/INFO->  [NAS]->  MOSS
3RAS->      RAS->     [RAS]->        RELVEC->    [VEC]->  VEC
4DDTDB->    DTTDB->   [DTDB/ETDB]->  DTDB/ETDB-> [ETDB]-> ETDB
```

[1]Data Generation System:
TerraBase - TerraBase Military Terrain Analysis System
DTEP-A/I - Digital Topographic Enhancement Program System - ARC/INFO
DTSS-P - Digital Topographic Support System - Prototype

[2]Data Format:
DIGIDATA - TerraBase (Interim)
DIGITAS - TerraBase (Final)
[CAS] - A/I Import: Compressed ASCII
ARC/INFO - DTEP-A/I
[NAS] - A/I Export: NonCompressed ASCII
MOSS - DTSS-P

[3]Data Structure:
RAS - RASTER
VEC - VECTOR
RELVEC - RELATIONAL VECTOR

[4]Data Feature Attribute Schema:
DDTDB - DIGIDATA Topographic Data Base
DTTDB - DIGITAS Topographic Data Base
DTDB - DTEP-A/I Topographic Data Base
ETDB - ETL Topographic Data Base

of feature data. Similar "attribute equivalence" identification techniques, albeit performed in a much more informal manner, were used for systematic examination of relationships and compatibility of attributes and associated values from within each of the various data bases selected for use within DTEP. Evaluation and depiction of such relationships between thematic layers of the data base was thus required in the initial portion of the effort. Individual "views" of feature encoding methodology were identified and data definition and content compared for consistency. Representative views of the Soil Type Category (STC) feature attribute and associated values are illustrated in Table 1.

## Table 1

**Attribute Analysis (Example)**

**TTADB Thematic Class: Surface Materials (Soils)**
**TTADB Attribute Field: Soil Type Category (STC)**

| Data Description | TTADB | TerraBase [1]DDTDB | | TerraBase [2]DTTDB | DTEP-A/I [3]DTDB | | | DTSS-P [4]ETDB |
|---|---|---|---|---|---|---|---|---|
| | - | F1 | F2 | F1 | F1 | F2 | F3 | F1 |
| Well-Graded Gravel | GW | 6 | - | 2 | 1 | 1 | 0 | A |
| Poorly-Graded Gravel | GP | 5 | - | 3 | 1 | 2 | 0 | B |
| Well-Graded Sand | SW | 12 | - | 4 | 1 | 5 | 0 | E |
| Poorly-Graded Sand | SP | 11 | - | 5 | 1 | 6 | 0 | F |
| Inorganic Clay Hi Plast | CH | 1 | - | 6 | 1 | 12 | 0 | M |
| Clayey Gravel | GC | 3 | - | 7 | 1 | 4 | 0 | D |
| Clayey Sand | SC | 9 | - | 8 | 1 | 8 | 0 | H |
| Inorganic Clay Lo Plast | CL | 2 | - | 9 | 1 | 10 | 0 | J |
| Silty Gravel | GM | 4 | - | 10 | 1 | 3 | 0 | C |
| Silty Sand | SM | 10 | - | 11 | 1 | 7 | 0 | G |
| Inorganic Silt Lo Plast | ML | 8 | - | 12 | 1 | 9 | 0 | I |
| | CL-MH | - | - | 13 | 1 | 17 | 0 | - |
| Inorganic Silt Elastic | MH | 7 | - | 14 | 1 | 13 | 0 | L |
| Organic Silt Lo Plastic | OL | 14 | 3 | 15 | 1 | 11 | 0 | K |
| Organic Clay Hi Plastic | OH | 14 | 2 | 16 | 1 | 14 | 0 | N |
| Peat /Hi Organic Soil | PT | 14 | 5 | 17 | 1 | 15 | 0 | O |
| Permanent Snow | PS | 14 | 4 | 18 | 1 | 16 | 98 | Q |
| Rock Outcrop | RK | 14 | 6 | 19 | 1 | 16 | 66 | P |
| Evaporites | EV | 14 | 1 | 20 | 1 | 16 | 24 | R |
| Open Water | W | 14 | 7 | 21 | 2 | 0 | 0 | W |
| BUA /Non-Evaluated Area | X | 13 | - | 1 | 1 | 18 | 0 | X |
| Unknown | - | 0 | - | 0 | 0 | 0 | 0 | Y |
| Void Collection Area | - | - | - | - | 3 | 0 | 0 | Y |
| Not Applicable | - | - | - | - | - | - | - | Z |

[1]DDTDB - DIGIDATA Topographic Data Base (TerraBase Data - Interim)

Raster data scanned from hard-copy TTADB factor overlays. Coordinate data and values for individual feature attributes are maintained within one or more binary files composed of a byte-screen dump with each byte a color code in the range of 0-15. Certain attribute information is stored in ancillary binary files of numeric data keyed to various combinations of feature attribute and related values. In the example above, appropriate combinations of attribute values for feature encoding of STC are located under subheadings F1 & F2. Two files are necessary for storage of STC attribute values as more than 16 unique soil type categories are required for proper data representation.

[2]DTTDB - DIGITAS Topographic Data Base (TerraBase Data - Final)

Raster data generated from DIGIDATA raster data. Coordinate data is maintained in a binary file of integers with pointers to a binary file containing unique cases or composite attribute values for 13 terrain feature attributes. Certain attribute information is stored in ancillary binary files of numeric data keyed to various combinations of feature attributes and associated values. In the example above, appropriate attribute values for feature encoding of STC are located under subheading F1.

[3]DTDB - DTEP-A/I Topographic Data Base

Relational vector data generated from DIGITAS raster data. Coordinate data is maintained in various binary files with pointers to relational files of tabular attributes and associated values maintained within an external data base management system (DBMS). Multiple fields of associated attribute information are maintained within one or more binary DBMS files. Individual values for multiple terrain feature attributes are stored as either alphanumeric characters or numeric data within one or more fields. In the example above, appropriate combinations of attribute values for feature encoding of STC are located under subheadings F1, F2 & F3 as several attribute fields are often necessary for proper data representation within the DTDB feature classification system.

[4]ETDB - ETL Topographic Data Base (DTSS-P Data)

Vector data generated from DTEP-A/I relational vector data. Coordinate data and values for individual feature attributes are maintained in a single binary file. Composite values for multiple terrain feature attributes are stored as alphanumeric characters within a single field. In the example above, appropriate attribute values for feature encoding of STC are located under subheading F1.

As anticipated, a fairly high degree of attribute equivalence was found between each of the various logical implementations of feature attribute encoding standards formulated for use in digital data creation from TTADBs and PTADBs. Primary factors of importance for attribute analysis and subsequent resolution of discrepancies in data description within the DTEP effort were found to include: the use, meaning or intent of an attribute or its encoded values; the set of values either enumerated by listing or specified as a range that an attribute might take; the unit of measurement applicable to an attribute and its associated values; and the set of allowable mathematical operations applicable to the values of a particular attribute. Minor inconsistencies in data definition were judged to be of little significance to overall data quality and accuracy. Discrepancies of consequence largely related to such factors as placement of data into arbitrary nominal ranges or categories rather than by storage of ordinal values obtained directly from the hard-copy source terrain feature overlays, variations in classification of such feature characteristics as "Unknown," "Not Applicable," "Non-Evaluated," "Void Collection Area," etc., and variations in compilation specifications used for creation of the original source hard-copy terrain feature overlays.

Formulation of specific technical DTD transformation procedures required an elaborate software development effort, including generation of various C programs and pc ARC/INFO macro language routines. Creation of necessary feature attribute conversion methodology was difficult to separate from that required for restructuring of the locational coordinate information, in large part due to diverse, interconnected data storage and indexing techniques in use for coordinate data and associated feature information. Procedures were developed to strip away the feature attributes from the source TerraBase coordinate data, reconfigure the attribute and coordinate files into appropriate structure and format for data transfer to DTEP-A/I, convert TerraBase feature attributes to those consistent with DTEP-A/I and DTSS-P specifications, import coordinate and feature attribute data into DTEP-A/I, relink DTEP-A/I and DTSS-P feature attribute files to corresponding coordinate data, and subsequently export data to DTSS-P structure and format with associated feature attributes.

Although a major hurdle involved the development of procedures to provide conversion of the physical structure and format of the TerraBase locational coordinate data to that compatible with DTEP-A/I (ARC/INFO) data import requirements, similar difficulties were confronted due to variations in physical storage characteristics of the nonlocational feature information. Much as the feature attribute classification system for a particular set of geographic information is based upon only one of many perceptions of the proper description or representation of such information, numerous techniques can be devised for feature file storage and linkage to associated locational coordinate data. Additionally, any particular attribute encoding system may have been formulated within the structural constraints of a proprietary data base management system (DBMS) or other unique data file structure and format. Such diversity in feature file storage methodology was a primary obstacle within the DTEP effort; e.g., ASCII format versus binary format, single field versus multiple field attributes, alphanumeric character versus integer, decimal or other numeric forms of data representation, single file versus multiple file structure and internal versus external linkage to coordinate data. Disparities in completeness and consistency of feature data encoding specifications and documentation, discrepancies between feature attribute information found within the digital data and that portrayed on the original source hard-copy terrain overlays, and a lack of available tools and utilities for identification and resolution of inconsistencies between feature attributes among data bases each also were found to be of significance to the effort.

DTEP - Phase 2

Successful implementation of spatial data exchange capabilities within DTEP has provided the impetus for continued development of spatial data creation and transformation procedures at USAETL. The scope of the effort has been expanded to include development of more extensive data generation capabilities from hard-copy TTADBs and PTADBs as well as the implementation of additional automated procedures for spatial data transformation and feature attribute conversion between other iterations of DTD. Continued development of such techniques will further the goal toward compatibility among data generated under evolving DTD standards for use in the DTSS and other systems.

## A LOOK TO THE FUTURE

Spatial data exchange capabilities are readily available for transformation of data structure and format between numerous GISs, as well as for import and export of various existing "standard" spatial data among many of these same systems. However, such data exchange capabilities which have been developed to date are generally limited to those necessary for transformation from one physical structure and format to another. Appropriate analytical techniques for identification of similarities and discrepancies in spatial data description and subsequent reconciliation of such inconsistencies from within any two or more reasonably compatible spatial data bases are relatively nonexistent.

Enhancements to current tools and utilities for use in the data transformation process could be incorporated into existing data import and export capabilities of various GISs. Alternatively, spatial data transformation software independent of the constraints of any particular system could be developed. Such methodology might include the capability for guiding the user through various menus to select appropriate spatial data transformation parameters. Parameters of significance which might be incorporated into such a package include structure and format of coordinate data and associated features; global information, including coordinate reference system, scale, geographic areal extent and other accuracy, quality and historical information; and attribute equivalence lookup tables necessary for conversion of feature attributes and related values from one classification system to another.

An immense volume of spatial data holdings constructed using a diversity of evolving systems, techniques and specifications are currently available; thus the demand for appropriate capabilities to more effectively use existing digital geographic information will most certainly continue to increase. Future research must address not only the requirements for structural transformation of data from one system to another, but conversion of attributes and related values to those compatible with other existing or newly produced data generated with alternate feature classification criteria. Assuming that content, accuracy and currency of existing spatial data lie within acceptable tolerances, enhanced data transformation capabilities incorporating various automated feature attribute conversion techniques might provide an effective means of support for revision of existing digital geographic information and integration of diverse spatial data.

## REFERENCES


Dangermond, J. 1988, GIS Trends and Comments: ARC News, Summer/Fall 1988, pp. 13-17.

Digital Cartographic Data Standards Task Force (DCDSTF) 1988, The Proposed Standard for Digital Cartographic Data: The American Cartographer, Vol. 15 No. 1, pp. 9-144.

Herring, J.R. 1988, A Fully Integrated Geographic Information System: Proceedings of the Ninth International Symposium on Computer-Assisted Cartography, pp. 828-837.

Marble, D.F. 1988, Approaches to Efficient Design of Spatial Databases at a Global Scale: in Building Databases for Global Science, edited by H. Mounsey and R.F. Tomlinson, Taylor & Francis, Ltd., Basingstoke, Hampshire, UK, pp. 49-65.

Nyerges, T.L. 1989, Schema Integration Analysis for the Development of GIS Databases: International Journal of Geographical Information Systems, Vol. 3 No. 2, pp. 153-183.

Pascoe, R.T and J.P. Penny 1990, Construction of Interfaces for the Exchange of Geographic Data: International Journal of Geographical Information Systems, Vol. 4 No. 2, pp. 147-156.

Rugg, R.D. 1988, Feature Definition: Methodology Used by Working Group III of the National Committee for Digital Cartographic Data Standards: Technical Papers of the 1988 ACSM-ASPRS Annual Convention, Vol. 5, pp. 108-117.

Wood, B.D. and D.H. Douglas 1983, Cartographic Feature Coding: Proceedings of the Sixth Annual International Symposium on Automated Cartography, Vol. 2, pp. 616-626.

## SPECIAL NOTE TO U.S. GOVERNMENT EMPLOYEES

**(To be completed by all authors and co-authors employed by the U.S. Government)**

Material prepared by U.S. Government employees as part of their official duties need not have the assignment of copyright transferred since this material is automatically considered as part of the public domain. If your paper falls within this category, please check the appropriate statement and sign below.

_x_ This paper has been prepared wholly as part of my (our) official duties as (a) U.S. Government employee(s). I (we) acknowledge that this paper has not previously been published, nor is it currently being considered for publication, by any other organization.

___ This paper has been prepared partly in the course of my (our) official duties as (a) U.S. Government employee(s). For any part(s) not prepared in the course of my (our) official duties, copyright is hereby transferred to ASPRS and ACSM. I (we) acknowledge that this paper has not previously been published, nor is it currently being considered for publication, by any other organization.

PAPER TITLE: Spatial Data Transformation: Feature Attribute Conversion Issues and Practical Experience

AUTHOR(S): Mark A. Sither

**Name(s) and Address(es)**

U.S. Army Engineer Topographic Laboratories

Telegraph and Leaf Roads, Bldg. 2592, Fort Belvoir, VA 22060-5546

SIGNED AND DATED: [signature] 1/15/91

**Author** **Co-author**

**Co-author** **Co-author**

**Co-author** **Co-author**

# STATEMENT OF CLEARANCE
(To be completed by all authors and co-authors)

If your paper is the result of work performed for the U.S. Government, you need a government public release approval. Please provide the following information:

PAPER TITLE: Spatial Data Transformation: Feature Attribute Conversion Issues & Practical Experience

AUTHORS: Mark A. Sitber

Name(s) and Address(es)

U.S. Army Engineer Topographic Laboratories

Telegraph & Leaf Roads, Bldg. 2592

Fort Belvoir, VA 22060-5546

Is any part of your paper the result of work performed for the U.S. Government? X Yes ____ No

If yes, please give the following information about your government public release approval:

Date of Approval ____________

Government office issuing approval U.S. Army Engineer Topographic Laboratories Liaison Office (CEETL-LO), Fort Belvoir, VA 22060-5546

SIGNED AND DATED: [signature] 1/15/91

Author

Co-author

Co-author

Co-author

Co-author

Co-author