# METADATA STANDARDS FOR FORESTRY AND NATURAL RESOURCES

M. H. Pelkki and G. J. Arthaud[1]

ABSTRACT.—Metdata are, in their most basic sense, data about data. Metadata compose the descriptions about data which make that data useful. This paper describes the need for metadata in both spatial and attribute databases. Current metadata standards are presented and their effects on forestry and natural resource information systems are discussed. The development of additional forestry and natural resource related metadata standards are then discussed with particular emphasis on incorporation of variance/error metadata, temporal metadata, and metadata to describe processes as well as states of the real world.

## INTRODUCTION: WHAT ARE METADATA?

Metadata are, in the simplest sense, data about data (Everest, 1986). They describe data in a database, defining content, scale, units of measurement, defining coded values, providing quality control information and explaining what purposes the data are used. In a very real sense, metadata make data useful. For example, in a forest inventory database, an attribute of a forest stand might be called BASAL AREA. In one instance of a stand, the value recorded for this attribute is "117." Metadata are needed to determine if the basal area is in square feet per acre, square meters per hectare, if the units are whole numbers of tenths, if the basal area pertains to only over story trees or only to woody stems with a particular diameter limit. In fact, the data item itself is small relative to the metadata needed to correctly use the data. Ganter (1993) states that metadata answer two fundamental questions about data:

(1) What are the data and do they meet our needs?
(2) How have the data been changed and what new data have been derived?

Ganter's first question covers the traditional, primary function of metadata, describing the current status of the data. But the second question moves metadata beyond a single snapshot of the data and into the realm of temporal data management and the concepts of raw or source data and derived or processed data.

Metadata are not only important for using data correctly, but it is also important when data is to be shared with multiple users. Curtice (1986) stated that the promise of databases is their ability to share timely and consistent data and to leverage existing data collection and storage against unanticipated, future requests for data. But all too often data do not conform to any standards and the semantics of data are buried in the database management system schema, application programs, users' and developers' minds (Boulanger and March 1989). This has led to growing stockpiles of stand alone databases with inconsistent data, single purpose projects and databases that waste information management resources and result in inconsistent analyses (Hofmeyr et al 1991). It leads to inconsistent application of data a users apply different definitions and standards to data at their choosing. And it also means that important metadata leaves the system when database developers are no longer available to provide metadata unrecorded by the system.

[1] Matthew H. Pelkki, Assistant Professor, Department of Forestry, University of Kentucky, Lexington, KY 40546-0073 and Greg J. Arthaud, Assistant Professor, School of Forest Resources, University of Georgia, Athens, GA 30602.

Obviously, metadata are important and though not a lot of thought has been given to metadata, its presence and frequency are increasing as more databases are viewed as shared corporate assets. As metadata become an integral part of databases, standards are needed to ensure a consistent, minimal set of metadata knowledge is maintained for all databases. Everest (1986) suggests that metadata should be available and managed via the same database management systems as the actual data. Any data not maintained by the database management system must be maintained by humans and as stated above, this leads to unsatisfactory data quality. Developing standards for spatial metadata has been a focus of the Federal Geographic Data Committee (FGDC) for several years in the 1990's (Tosta 1992, Tosta 1994). And while individual corporations and public agencies have data and metadata recommendations or standards, few spatial databases conform to these standards. And national standards for attribute data in forestry and natural resources have yet to be addressed.

The remainder of this paper will present three levels or general types of metadata. These metadata classes can be defined as directory-level, quality and accuracy report, and data dictionary. Examples of these will be provided below. Commonalities among these metadata will be noted and minimal metadata standards for forestry will be suggested.

## EXAMPLES OF FORESTRY METADATA

Two examples of directory-level metadata will be presented. The first is a directory of databases published by the Kentucky Cabinet for Natural Resources and Environmental Protection (Carey 1996). The second is metadata required by the FGDC (University of Texas 1995). Data-dictionary level metadata for the Eastwide (Hansen et al 1992) and Westwide (Woudenberg and Farrenkopt 1995) Forest Inventory Databases and FGDC-compliant metadata standards will conclude the presentation of examples.

### Summary of Data Bases: Natural Resources and the Environment in Kentucky

Kentucky State Regulations (KRS 61.940) requires that the state provide and make accessible an index of all state databases. From this, the Summary of Data Bases: Natural Resources and the Environment in Kentucky was published and distributed in 1996. This paper index is a directory of databases with each metadata field being a variable-length text field. Table 1 lists the metadata schema for this database. None of the meta-data attributes are defined, and the description of each databases' attributes is limited to their system name, with no definition of individual attributes. The only possible quality attribute is "Integrity," and while it is listed as either "Good" or "Excellent," the meaning is never defined. Also, there are no temporal metadata. Knowing when the database was created, last updated, or the frequency of updates would be very useful. While this is very rudimentary metadata, it is still very useful.

### Minimum metadata required by FGDC Spatial Metadata Standard

The FGDC has developed national standards for spatial metadata to assist in data sharing and coordination. There are 10 main sections of FGDC metadata:

1. Identification Information
2. Data Quality Information
3. Spatial Data Organization Information
4. Spatial Reference Information
5. Entity and Attribute Information
6. Distribution Information
7. Metadata Reference Information
8. Citation Information
9. Time Period of Content Information
10. Contact Information

From this list, the directory-level information can be found in items 1, 7, 8, 9, and 10, above. A detailed listing of the minimum directory-level metadata is shown in Tables 2 and 3. Note that this metadata provides for changing metadata standards and is quite detailed compared to the Summary of Kentucky Natural Resource and Environmental Databases.

**Metadata provided by Eastwide and Westwide Forest Inventory Databases**

Perhaps the most widely used forest inventory databases are the Eastwide and Westwide Forest Inventory databases developed by the USDA Forest Service Forest Inventory and Analysis Unit. The documentation for these two databases (Hansen et al 1992, Woudenberg and Farrenkopt 1995) provides data dictionary-level metadata. However, in the narrative, general quality level information is provided when the data is used on a statewide basis and cautionary statements are provided concerning data quality when subdividing the data into smaller geographic units. Table 4 provides a listing of the metadata recorded for these databases.

This is fairly typical of metadata for published databases and is really the minimum metadata for sharing the data. However, it tells us little to nothing about the data. Individual measures of data quality are absent, as are acceptable ranges for data values. The data-dictionary metadata should provide users and managers with information that helps them manage the database by identifying values that are out of range or incorrectly measured, and it should give the user some idea of data quality.

**Table 1. Metadata for Kentucky Natural Resource and Environmental Databases**

| | | |
|---|---|---|
| 1. Name of database | 8. Contact name | 15. Descriptors/Keywords |
| 2. System name of database | 9. Contact address | 16. Parameters (attributes) |
| 3. Purpose of database | 10. Contact phone | 17. Forms |
| 4. Agency of administration | 11. Host computer | 18. Sources |
| 5. Department | 12. Language | 19. Comments |
| 6. Division | 13. Coverage | |
| 7. Branch | 14. Integrity | |

1 Citation:
    Citation_Information:
    Originator:
    Publication_Date:
    Title:
2 Description:
    Abstract:
    Purpose:
3 Time_Period_of_Content:
    Time_Period_Information:
    Single_Date/Time:
    Calendar_Date:
    Currentness_Reference:
4 Status:
    Progress:
    Maintenance_and_Update_Frequency:
5 Spatial_Domain:
    Bounding_Coordinates:
6 Keywords:
    Theme:
    Theme_Keyword_Thesaurus:
    Theme_Keyword:
7 Access_Constraints:
8 Use_Constraints:

Metadata_Date:
Metadata_Contact:
    Contact_Information:
    Contact_Organization_Primary:
    Contact_Organization:
    Contact_Address:
    Address_Type:
    Address:
    City:
    State_or_Province:
    Postal_Code:
    Contact_Voice_Telephone:

Metadata_Standard_Name:
Metadata_Standard_Version:

1 Structure of record (County, Plot, or Tree)
2 Attribute Name
3 Computer alias
4 Column position
5 Format
6 Unit of measure
7 Frequency/Restriction (attribute required or not)
8 Key field (yes or no)

**Data-dictionary metadata required by the FGDC Spatial Metadata Standard**

Contrast the data-dictionary for the FIA data (Table 4) with that required for entities and attributes in the FGDC Spatial Metadata standard (Table 5). Information on attribute domain (the valid set of values for an attribute) is greatly expanded, and there is provision for data quality as well. Source information is also provided for definitions and codesets and providing easier linkages among datasets by identifying sources of definitions for metadata. And also important are the temporal meta-data, which provide the user with data currency information.

- 1 Detailed Description
  - Entity type
    - Entity type label
    - Entity type definition
    - Entity type definition source
  - Attribute type
    - Attribute label
    - Attribute definition
    - Attribute definition source
    - Attribute domain values
      - Enumerated domain
        - Enumerated domain values
        - Enumerated domain value definitions
        - Enumerated domain value definition source
      - Range domain
        - Range domain minimum
        - Range domain maximum
      - Codeset domain
        - Codeset name
        - Codeset source
      - Unrepresentable domain
    - Attribute units of measurement
    - Attribute measurement resolution
    - Beginning date of attribute values
    - Ending data of attribute values
    - Attribute value accuracy information
      - Attribute value accuracy
      - Attribute value accuracy explanation
    - Attribute measurement frequency
- 2 Overview description
  - Entity and attribute overview
  - Entity and attribute detail citation

## WHERE DO WE GO FROM HERE?

We rely daily upon all kinds of information system standards. Hardware and software standards allow us to run operating systems and application software on numerous machines with high degrees of confidence. Processing and application standards allow data to be shared and we can be assured of consistent results. Likewise, metadata standards will improve our ability to exchange data, provide relevant and current data, and enhance our data access. The problem with metadata isn't that researchers and practitioners don't find it useful. The problem with metadata is that, like documenting software, it takes time and effort to do it right! Tables 2, 3, and 5 portray only a portion of the FGDC Spatial Metadata Standard and this metadata alone would pose a significant burden on database developers. In some cases, the metadata could require more storage and management effort than the data themselves! To manage this problem, the following metadata management suggestions are provided:

**Maintain directory-level metadata**. Research units, companies, government agencies should maintain a directory of databases and for those databases the metadata should describe those

attributes found in table 1 and also metadata concerning date of collection, a temporal range of when the data are valid descriptions of the real world, how frequently the data are used, and some global indication of data accuracy and quality. Organizations that know what data they have, who manages it, and how to access it can share data more easily within the organization and can provide data-directories to cooperating organizations as well.

**All metadata should be maintained using a database management system**. Many directory level databases are maintained on word processors rather than on database management systems (DBMS). The metadata should be maintained on the same DBMS as the actual data, as many DBMS can use metadata to assist in managing data. For example, metadata concerning valid domains for a data field are used to error-check a database as well as inform potential users about the range of values found in the reality represented by the database.

**Identify which databases are truly corporate and expand their metadata**. Almost everyone builds databases. But only those databases which are shared or have great potential for sharing are worth the effort of expanding from directory-level to data dictionary-level. How do you identify such databases? Some features may include 1) a large number of requests for use by second or third parties, 2) cost of development (expensive data is worth documenting), 3) wide geographic extent, and, 4) generalized nature of data (databases that support many functions within an organization).

**Build from directory-level to data-dictionary level metadata**. It is likely that data-dictionaries exist, either explicitly or buried in application programs, DBMS attribute definitions, or users and developers minds. Established directory level metadata will increase the reuse of databases and build impetus for creating and maintaining detailed on-line data dictionaries.

The name of the game with metadata is leveraging data assets by assisting users in finding data, assuring them of the currency and quality of that data, and providing means of integrating data from various sources into a new database that is used for purposes unforeseen by the original database developers.

## LITERATURE CITED


BOULANGER, D., AND S.T. MARCH. 1989. An Approach to Analyzing the Information Content of Existing Databases. *DATA BASE*, Summer 1989, pp. 1-8.

CAREY, D. 1996. *Summary of Data Bases: Natural Resources and the Environment in Kentucky*. Kentucky Geological Survey, University of Kentucky, Lexington, KY 40506. 186 p.

CURTICE, R.M. 1986. Getting the Database Right. *Datamation*, Oct. 1, 1986. pp 99-104.

EVEREST, G.C. 1986. *Database Management*. Objectives, System Functions, and Administration. McGraw-Hill, New York, NY. 816 p.

GANTER, J.H. 1993. Metadata management in an environmental GIS for multidisciplinary users. *GIS/LIS /93 Proceedings*. Minneapolis, MN, U.S.A. ACSM-ASPRS-URISA-AM/FM, 1993.1: 233-245.

HANSEN, M.H., T. FRIESWYK, J.F. GLOVER, AND J.F. KELLY. 1992. *The Eastwide Forest Inventory Data Base: Users Manual*. U.S. Department of Agriculture, Forest Service. Gen. Tech. Rep. NC-151. 48 p.

HOFMEYR, H.D., D.E. ALLEN, AND E. KENK. 1991. Setting the Stage for Data Sharing in British Columbia. *GIS*

*'91 Symposium*, Vancouver, British Columbia, Feb. 1991. pp. 89-93.

TOSTA, N. 1992. Data Data - The National Spatial Data Infrastructure. *GeoInfo Systems* 2(10):30-35.

TOSTA, N. 1994. Data Data - The National Spatial Data Infrastructure: Where Are We Now? *GeoInfo Systems* 4(1):25-28.

UNIVERSITY OF TEXAS, BRUTON CENTER, SCHOOL FOR SOCIAL STUDIES. 1995. *Minimum for FGDC Compliant Data*. Internet Web Page: http://www.bruton.utdallas.edu/