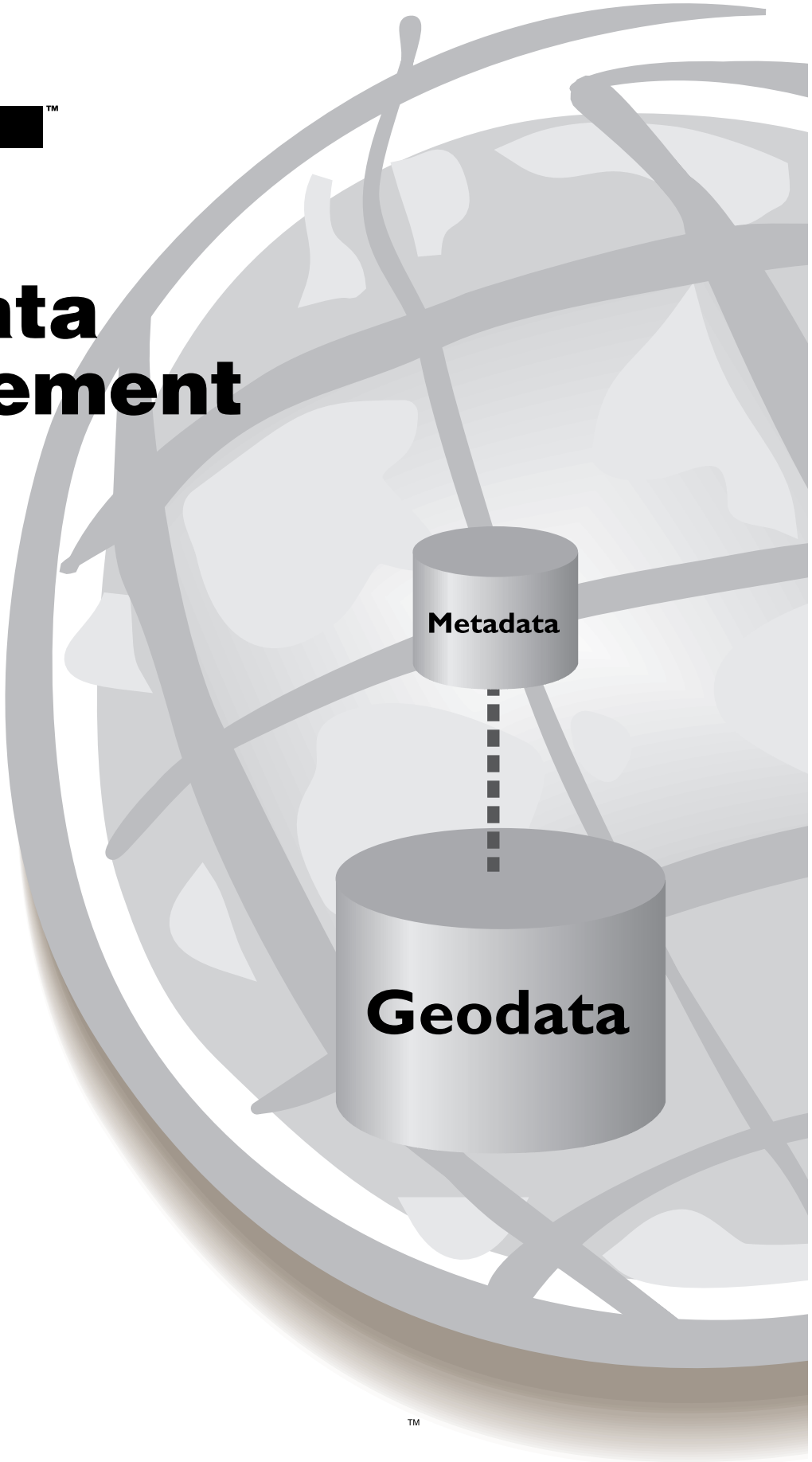
GIS by ESRI™

Metadata Management in GIS



ESRI®
White Paper Series
August 1995

ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE, INC.

Copyright © 1992–1995 Environmental Systems Research Institute, Inc.
All rights reserved.
Printed in the United States of America.

The information contained in this document is the exclusive property of Environmental Systems Research Institute, Inc. This work is protected under United States copyright law and other international copyright treaties and conventions. ESRI grants the recipient of the ESRI information contained herein the right to freely reproduce, redistribute, rebroadcast, and/or retransmit this information for personal, noncommercial purposes, including teaching, classroom use, scholarship, and/or research, subject to the fair use rights enumerated in Sections 107 and 108 of the Copyright Act (Title 17 of the United States Code). No part of this work may be reproduced or transmitted for commercial purposes in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage or retrieval system, except as expressly permitted in writing by Environmental Systems Research Institute, Inc. All requests should be sent to Environmental Systems Research Institute, Inc., 380 New York Street, Redlands, CA 92373 USA, Attention: Contracts Manager.

The information contained in this document is subject to change without notice.

RESTRICTED RIGHTS LEGEND
Use, duplication, and disclosure by the government are subject to restrictions as set forth in FAR §52.227-14 Alternate III (g)(3) (JUN 1987), FAR §52.227-19 (JUN 1987), or DFARS §252.227-7013 (c)(1)(ii) (OCT 1988), as applicable. Contractor/Manufacturer is Environmental Systems Research Institute, Inc., 380 New York Street, Redlands, CA 92373 USA.

ESRI, ARC/INFO, PC ARC/INFO, ArcView, and ArcCAD are registered trademarks; ARC COGO, ARC NETWORK, ARC TIN, ARC GRID, ARC/INFO LIBRARIAN, ARCPLOT, ARCEDIT, TABLES, Application Development Framework (ADF), ARC Macro Language (AML), Avenue, FormEdit, ArcSdl, ArcBrowser, ArcDoc, ARCLine, ARCSHELL, IMAGE INTEGRATOR, DATABASE INTEGRATOR, DBI Kit, ARC/INFO, ArcTools, ArcStorm, ArcScan, ArcExpress, ArcPress, Mapplets, SPATIAL DATABASE ENGINE (SDE), PC ARCEDIT, PC ARCPLOT, PC DATA CONVERSION, PC NETWORK, PC OVERLAY, PC STARTER KIT, PC ARCSHELL, Simple Macro Language (SML), ArcUSA, ArcWorld, ArcScene, ArcCensus, ArcCity, the ESRI corporate logo, the ESRI globe logo, the ARC/INFO logo, the ARC COGO logo, the ARC NETWORK logo, the ARC TIN logo, the ARC GRID logo, the ARCPLOT logo, the ARCEDIT logo, the Avenue logo, the ArcTools logo, the ArcStorm logo, the ArcScan logo, the ArcExpress logo, the PC ARC/INFO logo, the ArcView logo, the ArcCAD logo, the ArcData logo, the SDE logo, the ArcPress Logo, the GISTOOLS logo, ARCware, *ARC News*, ArcSchool, ESRI—Team GIS, ESRI—The GIS People, GIS by ESRI, ARC/INFO—The World's GIS, Geographic User Interface (GUI), Geographic User System (GUS), Geographic Table of Contents (GTC), Your Personal Geographic Information System, the World's GIS, GIS for Everyone, and The World's Leading Desktop GIS are trademarks; and the ArcData Publishing Program, ARCMAIL, ArcQuest, ArcWeb, ArcOpen, www.esri.com, @esri.com, and Rent-a-Tech are service marks of Environmental Systems Research Institute, Inc.

The names of other companies and products herein are trademarks or registered trademarks of their respective trademark owners.

J-7056

Metadata Management in GIS

An ESRI White Paper

Contents **Page**

Introduction 1

What Is Spatial Metadata? 2

Why Is Spatial Metadata Important? 4

Standardizing Spatial Metadata 6

The Content Standards for Digital Geospatial Metadata 7

How Much Spatial Metadata Is Enough? 9

The FGDC Standard and the SDTS 11

International Spatial Data Standards 12

On-line Publication of Spatial Metadata 13

Environmental Systems Research Institute, Inc. (909) 793-2853
380 New York Street, Redlands, CA 92373 Fax (909) 793-5953

Current Metadata Support in ARC/INFO 15

Metadata Generated and Maintained by ARC/INFO 15

ARC Macro Language-Based Metadata Management 15

Metadata Capture with DOCUMENT.AML in ARC/INFO Version 7.0.3 16

Future Metadata Management for ESRI Products 17

Metadata Management Functionality 18

Adding New Data Types to the ESRI Data Model 19

Future Software Product Plans 19

Summary and Conclusion 20

Suggested Readings 21

J-7056

Metadata Management in GIS

An ESRI White Paper

Introduction

A number of recent technological advances are changing the role of geographic information systems (GISs) in many organizations. Industry trends are shifting from a GIS-centric view to an information system (IS) department-centric view with GIS as an integrated component, and from integrated applications to component-based applications where corporate data structures dominate and geodata are integrated into these structures. A comparable trend is visible in the area of spatial databases: numerous organizations have matured from building to using their geographic databases as well as making them publicly available to others. This trend has led to an explosion of available geodata sets, which, along with the recent advances in network computing (more commonly known as the information superhighway), is helping to accelerate spatial data sharing. Coupled with these trends, organizations have recognized the need to manage their data holdings to remain productive and competitive. Spatial metadata provide a viable solution to this data management problem.

This white paper describes the Environmental Systems Research Institute, Inc. (ESRI), view on what metadata are, why metadata are important, why spatial data standards are important, and what standards do for you. It provides details on currently existing metadata implementation in ARC/INFO® software and introduces the metadata managing functionality under development for ARC/INFO Version 8.0. Issues relating to spatial metadata publication are briefly addressed, followed by concluding remarks and a list of suggested readings.

Environmental Systems Research Institute, Inc. (909) 793-2853
380 New York Street, Redlands, CA 92373 Fax (909) 793-5953

What Is Spatial Metadata?

In answering this question, we take the real world as the source for all our data. The collected data describe any feature or subject you are interested in about the real world on a more general level. Take, for example, the Wyndham Hotel, the site of ESRI's annual user conference. The Wyndham is a physical, real-world object. Data stored about it in the City of Palm Springs' GIS database include its parcel boundary and building footprint, with precise location reference as the graphic representation. Also stored are descriptive data—in a tabular database—that describe the Wyndham's characteristics. Examples of such descriptive data are lot number, lot size, lot owner, building owner, building material, year built, assessed value, number of stories, number of guest rooms, restaurants, meeting rooms, list of staff, and so on.

Metadata describe these data about the Wyndham in yet more general and less detailed terms. Thus, metadata are frequently referred to as data about data. Data about the Wyndham are then grouped into the following metadata categories: geographic location for parcel and for building, property record, construction record (including metadata images such as CAD blueprints and photos), and usage record with the subcategories room types, dining facilities, recreational facilities, conference facilities, employee record, and others.

Another set of metadata documents bibliographic information such as who collected the data on the Wyndham; when they were collected; how they were collected, preprocessed, and converted; their resolution; who holds the data now; and so on. Metadata are, therefore, also referred to as additional information that is needed for the data to be useful (i.e., information that facilitates understanding between the provider and the user of the data and their content).

A point worth mentioning here is made by F. Bretherton (1994, p. 3), who states that "there is no LOGICAL distinction between metadata and data"; rather, it is the database context and the user's degree of understanding of the database content that determine if metadata are data or data are metadata. Clearly, the term means different things to different people in different contexts at different times. The term defies a single and precise definition and most certainly goes beyond the mere recording of existing data sets.

Metadata are, then, comprehensive, systematic, and deductive information about the content, the structure, the relationships, the representation, and the use context of the data stored in the underlying database. Metadata management is multidimensional; it includes the following aspects:

- Inventorying existing data holdings
- Defining the names and data items to facilitate understanding, a common basis for communication, and sharing across user domains
- Building a keyword list of names and definitions
- Indexing the inventory and the keyword list for access
- Recording processing steps performed on the data including those involved in the initial collection
- Documenting the data structures used and the data model implemented
- Recording the logical and the physical database schema
- Documenting the relationships between data items in different data sets as well as between individual data sets
- Recording the processing steps performed on the data including precollection decisions, collection methods, conversion, and postconversion editing and analyses
- Documenting the representation chosen for the data
- Documenting application-specific metadata including flowcharts for macro language programs
- Updating the metadatabase in a consistent fashion and at regular intervals

This white paper focuses on the metadata documenting existing data sets for publication and addresses feature class metadata as they apply to an ARC/INFO project.

Why Is Spatial Metadata Important?

Spatial metadata are important because they document existing data holdings and facilitate data sharing. Metadata are an essential resource that you share with many users at all levels of GIS expertise. Both the demand for and supply of spatial data continue to grow rapidly, while costs continue to remain high. Given these realities and the fact that the acquisition of data for a project is critical to its success, it is the economically sensible thing to do that you avoid duplication of effort and use existing spatial data as well as share your own. This requires you to catalog and adequately describe your spatial data sets and make this information publicly available for your own and others' benefits, both inside and outside your organization. Submitting to this discipline benefits you in at least two ways: you gain understanding on what to look and ask for in your own searches and you come to appreciate the work others have put in to publish their geodata sets. While it is time-consuming and demands commitment, the benefits you gain from capturing spatial metadata clearly outweigh the work involved.

Spatial metadata are important because they reduce the volume of typically very large spatial data sets to a searchable, while meaningful, size. Another objective of capturing metadata is to provide a summary view of the essential characteristics of spatial data sets, which allows the user to make an informed assessment of their fitness for use for a specific application. It is, therefore, highly desirable and requires thought and discipline to ensure that the volume of the metadata does not exceed the volume of the data set they describe. Note, on the one hand, the capture of metadata increases the overall size of your database; on the other hand, your search questions are processed against the metadata portion, which means less data to process and a faster return of an answer. Adequate, easy-to-use search, browse, and retrieval mechanisms for spatial metadata support your speedy GIS application development.

Spatial metadata are important because they provide information on the data otherwise not readily available. This proposition answers the question: Why collect metadata? Think of metadata as information you need to know in order to use the data. Metadata represent a set of characteristics about the data that are normally not contained within the data themselves and that quite frequently only exist in the memory of the persons involved in the data's collection and use. Take, for instance, information about the quality of a data set: You can obtain such information only at the time you purchase the data. Clearly, on-line metadata, including a brief sample and a browse image, are the most effective means to tell you if a data set matches your project's needs. It allows you to evaluate the benefits and shortfalls involved prior to purchasing and using a data set.

Spatial metadata are important because they support software-based and organizationwide standards. The benefit of having software-based data standards is that the program is easier to use, and users can readily move data between systems and platforms. A benefit not to be overlooked in the organizational environment is that collecting and storing spatial metadata help you to establish and enforce naming, definition, cataloging, and operating standards for all departments; this in turn is the vital foundation for you to understand, collaborate, and share your resources and use those of others.

Spatial metadata are important because they support easier spatial data access and management. Metadata provide a guide to the casual and novice user's question, "How do I know what to ask for?" with a synoptic view on what is available on their area of interest, where the information is, how current it is, what format it is in, and what use constraints apply. For spatial data professionals, for instance, a map production company—which typically captures and uses rather detailed metadata—metadata provide feature- and attribute item-level metadata management. Frequent updates are readily accommodated and these updates are easily integrated into the daily production process. Metadata documentation is thus not an end in itself; rather, you can view it as a tool that will greatly improve your work with spatial data and increase your overall GIS benefits.

Standardizing Spatial Metadata

Why are standards so important? Standards establish a common ground on which to work and build. As such, they facilitate understanding and communication between people from different backgrounds and different levels of expertise. Take, for example, road signs: the shape of the stop and the yield sign are universal. (Confusion and misunderstanding are primarily based on the ambiguous use of words.) Standards ensure that you yield a consistent and uniform product and, based on standardized input and standardized operations on that input, they allow you to evaluate the quality of the output product.

It is not an easy task to develop standards, particularly in the GIS field, which brings together people from many different disciplines/ environments with varying degrees of sophistication. Yet, our diverse backgrounds is exactly the reason why we need standard names and definitions and a standard to document the content of our spatial data. Standards are intended to make things easier and help us communicate, collaborate, and benefit from the power GIS offers. They facilitate the sharing of spatial data and spatial data products within and across organizations and across hardware and software environments.

Establishing standards means spatial data users and producers must agree on and use standard terminology and definitions, especially with respect to the spatial data meaning and their relationships. Surely, some of you have experienced the frustration and difficulties when working with spatial data that were collected without any standard or with a different standard then the one you are familiar with. Or worse yet, you used a data set that contained errors and you did not know about it. The benefits of a standard for spatial metadata are obvious: Metadata will provide you with the information you need to evaluate spatial data, while standard names, definitions, and content ensure that you understand someone else's metadata content and they understand yours.

Quality assessment is a particularly important aspect in the realm of digital spatial data, not only because spatial data are very expensive but also because their quality has potentially unforeseeably far-reaching consequences to your project and the decisions based on its

results. Standards and metadata allow you to evaluate the fitness for use of the geodata set you consider to purchase. If you know how the data were created and processed, you have a frame of reference by which to measure their quality and usability. Reliable data are the basic component of an effective GIS.

The Content Standards for Digital Geospatial Metadata

The need for digital spatial data standards has been recognized since the early days of GIS. However, standards issues are only addressed when an actual—and frequently urgent—need exists (i.e., when working without standards becomes a real nightmare). In the case of the Spatial Data Transfer Standard (SDTS), it was the need to send digital spatial data sets across a network to other platforms and software systems without loss of data (USGS, 1992). In the case of the Content Standards for Digital Geospatial Metadata (from hereon referred to as FGDC Standard), it was the (mandated) need of federal agencies and other organizations to document their spatial data holdings and make the documentation publicly available. This documentation is primarily concerned with so-called bibliographic metadata in mostly textual format.

Along with the evolving GIS technology and the tremendous increase in operational systems in the past five years, data standards evolved and have been implemented, among them the SDTS, which we address in more detail below. It was not until 1990 that the Federal Geographic Data Committee (FGDC) was formed and, among other responsibilities, was assigned the task to develop a content standard for digital spatial metadata. The committee consists of policy-level representatives of all U.S. federal agencies and is an integral part of the National Spatial Data Infrastructure (NSDI). NSDI has been proclaimed by Executive Order 12906 ("Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure," signed by President Clinton on April 11, 1994) and requires all federal agencies to document their spatial data holdings and make spatial metadata accessible to the National Geospatial Data Clearinghouse. The Clearinghouse is a distributed network that allows public and private data providers to publish their spatial data with instructions on how to obtain the data.

Out of the five-year work of the FGDC evolved the FGDC Standard (FGDC, 1994). The FGDC Standard has been in effect for all federal agencies since January of 1995. Its purpose is to provide a consistent nomenclature on what kind of spatial metainformation to capture. On this basis, it facilitates searching for spatial data and evaluating their fitness for use over the network and through the Clearinghouse. How to capture, store, and manage its seven categories—they are nonhierarchical and nonordered—is intentionally not addressed by the FGDC Standard, allowing flexibility for its implementation. The seven categories are

- Identification Information
- Data Quality Information
- Spatial Data Organization Information
- Spatial Reference Information
- Entity and Attribute Information
- Distribution Information
- Metadata Reference Information

Most of the metadata entries in these categories are termed mandatory and mandatory where applicable. At close scrutiny, several sections termed mandatory where applicable are inconsistent and require you to carefully walk through a complete example to be able to determine for yourself what to leave out and what to fill. While only mandatory for the U.S. federal agencies, the FGDC hopes that the NSDI and its Clearinghouse furnish the incentive for the private sector to adopt the FGDC Standard, too, and contribute their spatial data to the Clearinghouse.

Metadata documentation currently underway at the private level shows that organizations continue to establish their own standards while orienting themselves with differing degrees on the FGDC Standard and tailoring it to their needs. Many of you who have developed a data dictionary to document basic organization-related metainformation look to extend these initial efforts with more comprehensive metadata documentation. Others of you, ready to get the process started, look to your software vendor for guidelines on how to implement the

FGDC Standard's extensive, and mostly mandatory, content. In this context, an often asked question is:

- How much spatial metadata is enough?

How Much Spatial Metadata Is Enough?

The answer to this question depends on which user questions your metadata must answer, which depends on the user's needs, which, in turn, depends on the project's scope and the individual project tasks. Clearly, there is no single answer to this question. This is why it is very important that you thoroughly analyze your GIS application needs, record them, and thoroughly design an extensible system before implementing a solution.

Metadata management is a difficult task. The difficulty lies in the very nature of the spatial data itself. Spatial data sets are typically very large in volume and contain a very large number of different object types, resulting in complex database schemata. Their heterogeneous nature further compounds the difficulty: Spatial data frequently come in different data formats such as vector (topological, nontopological), raster, triangulated irregular network (TIN), object-oriented, tabular, and text. They come from heterogeneous sources, each with its own terminology and associated definitions (or lack thereof). Spatial data vary greatly in spatial and temporal resolution and range widely in quality. This enormous diversity represents a formidable task and challenge for those setting out to document spatial data sets; yet, they must be documented if they are to be used by people other than their creator.

Published metadata project reports reveal that it is essential to stay focused on what is needed and choose a simple, step-by-step approach to this formidable task (Miller and Forner, 1994). Metadata records range from simple descriptions to highly structured and detailed listings of every possible metadata element at the feature and attribute item level, as is needed for a vehicle navigation database. Simple descriptions of metadata records include the following:

- What data exist?
  - Name
  - Theme
  - Brief description
  - Areal extent
- Are they of any use to me?
  - What are their contents (theme)?
  - Their resolution
    - Spatial
    - Descriptive
    - Temporal
  - Their format
    - Data format
    - Medium
  - Their age (creation date)
- Can I see a sample?
- Is this what I really want?
- How do I get them?
  - Contact
  - Custodian
  - Transfer price

An example of a structured detailed listing is the metadata for a road. This listing would include the road's location, year built, its size, its intersections and related objects, its condition, its surroundings, as well as who built the database and when, the database schema, and descriptive data coding schemata. The FGDC Standard, though only

minimally structured, is an example of a detailed documentary—though not a feature-level—spatial metadata inventory.

As you consider the options available to you when you plan your spatial data set documentation project, let your own questions guide you. A 1994 study in Australia (Miller & Forner, 1994; Miller & Bullock, 1994) has revealed that there exists a set of the ten most commonly used metadata elements. They are name, abstract, contact, data format, spatial extent, data currency, custodian, data form, contact keywords, and order procedure. These items translate into three primary user metadata questions: What are there?, Where are they?, and How do I get them? Start out with setting up a table and enter the most elementary and most commonly asked questions such as the ones just mentioned, plus What format are they in? and How old are they? and other questions you want to ask. Next, enter the data needed to answer these questions. A general rule of thumb is to ensure that your most general as well as your most structured and focused questions can be answered by the metadatabase.

The FGDC Standard and the SDTS

Both standards have distinctly separate, yet complementary, functions. The FGDC Standard is solely concerned with what metadata you need to collect about a spatial data set, while the SDTS's sole purpose is to ensure that these data and the data you import or export transfer one-to-one if differing computer platforms and software systems are involved. It also ensures integration of data with a different data model. ARC/INFO Version 7 provides full support for the SDTS. For more details on the SDTS and ARC/INFO refer to the ESRI white paper on the subject.

Metadata on the quality of the data set—known as the quality report, a hard-copy document—are an important and mandatory component of the SDTS and thus of every data set transfer. The other metadata are contained in a very basic data dictionary, describing standard spatial objects, descriptive attributes, and valid attribute domains and value ranges. The data quality and the data dictionary information are the only sections common to both standards. The FGDC Standard replicates the content (not the structure) of SDTS's data quality section, and the latter's data dictionary information is captured under

entity and attribute information in the former. Both standards require the data provider to extensively document data quality metadata, while entity and attribute metadata are limited and more rudimentary.

Currently, the biggest shortfalls of the SDTS are that its metadata are only available to you after the data transfer and that you cannot query them. Implementation of the FGDC Standard will help to change this. Its purpose is to make metadata available for you to search and query on-line via the NSDI's Clearinghouse prior to acquiring a spatial data set. Following your successful search and your decision to purchase data, the SDTS's protocol ensures a one-to-one transfer of these data to your system.

International Spatial Data Standards

The need for spatial data set documentation and metadata management is international. All countries around the globe involved in GIS are also involved in establishing standards for digital spatial data. To date, internationally the farthest progress has been made in the area of spatial data exchange standards.

Most central and northern European and several Middle Eastern countries have developed and instituted their own transfer standards. Among the most widely used are Britain's National Transfer Format (NTF), Germany's and Belgium's Geographic Data Files (GDF)—used in the development of the European Digital Road Map (EDRM) project—and the Digital Geographic Information Exchange Standard (DIGEST) developed by military and civilian mapping organizations from Belgium, Canada, France, Germany, Italy, the Netherlands, Norway, Spain, Great Britain, and the United States. Each European ESRI office has developed a translator for the respective standard's use in ARC/INFO. The current European efforts go toward unifying these initiatives, with GDF having very strong support from Centre European de Normalisation (CEN) to become the official European Geodata Standard (Ostyn, 1995). Australia has officially ratified the SDTS, while many south and Southeast Asian countries look to the Western world for guidelines and orientation.

To date, no details are available on metadata standard initiatives—comparable to the FGDC Standard—at the national level outside the

United States. In most European countries, among them Sweden, Finland, Norway, Britain, the Netherlands, Germany, France, and Belgium, spatial metadata documentation progresses on an individual project basis. The most widely known European project is the European Space Agency Prototype International Directory, an on-line geosciences metadata system (Walker, 1991). A Swedish project aims to set up a common marketplace and initiate spatial data sets exchange, similar to the FGDC's Clearinghouse. The Norwegian SAMPO project uses ArcView® Version 2.0 to catalog its spatial data holdings and provides dial-in access (Mikkonen and Rainio, 1995). Similarly, the Dutch Ministry of Transportation, Public Works and Water Management (Kuggeleijn, 1995), the German State Mapping Agencies' ATKIS (personal communications), the New South Wales Department of Conservation and Land Management's Data Directory (Miller and Forner, 1994), and the Australian FINDAR system (Johnson et al., 1991) projects are documenting their spatial data holdings to satisfy their own needs.

All these projects clearly show that metadata capture supports the development of spatial data standards and is required for sharing data via a spatial data clearinghouse.

On-line Publication of Spatial Metadata

One primary reason for collecting bibliographic metadata for spatial data sets is its use for on-line publication. Two factors are imperative when publishing your geodata: the use of standard terminology when describing your data set, which is what much of this white paper is all about, and regular updates and posting. If you follow the FGDC Standard, your spatial metadata provide standard field names holding information needed to search and evaluate spatial data over a network or through a clearinghouse. As the FGDC points out, we already use such standard names and text in the scale, legend, and accuracy information on every paper map to determine its usefulness for our purposes. If you have ever been in a situation where you realized that the data you acquired were out-of-date, the experience left you hopefully determined to ensure that your published metadata are up-to-date.

On-line publication of spatial metadata not only makes sense, given today's economic realities, it is the only way to go if you want to stay current and remain productive and competitive. Three Internet options are available to you: the Internet's specialized list servers; the Use net, an internationally circulated series of bulletin boards from which you select the ones you are interested in; and the World Wide Web (WWW, or the Web), a global network of multimedia hypertext servers, which includes Gopher as a popular application (Gittings, 1995). These servers provide you with access to the most up-to-date information.

The use of any of the Internet or Use net for publishing spatial metadata is rather straightforward, both producing excellent results if you deal well with mostly textual and frequently cryptic information. The Web organizes and presents information visually and in thematic units based on the central principle of hypermedia: associative, progressive disclosure. It uses HyperText Markup Language (HTML) text with typically one to many connected documents, which you can access by selecting a specially highlighted (i.e., HTML-tagged and -linked) keyword. The same mechanism applies to this newly opened document and all successive ones. As you progress in your search, a special function keeps track of your path and allows you to retrace it entirely or only partially.

The spatial metadata you and others have documented and posted on the Web (note that ESRI will provide you with the necessary functions to place your spatial metadata on the Web) can be navigated in exactly the same fashion, allowing you to either go deeper—accessing more detailed information—or branch out and explore adjacent areas, much like the pan and zoom operations in the drawing/mapping environment. This visual environment is very conducive to the use of the earlier mentioned browse image, which allows the searcher to see for her/himself what the data look like. On the Web, your search for spatial data is fun, informative, and hopefully productive.

Current Metadata Support in ARC/INFO

Spatial data documentation is no longer something you will eventually do if you get around to it. Rather, documenting geodata and with it metadata management have become an economical necessity. This becomes clear once you are aware of the metadata that currently exist and the many uses they have and offer you.

Metadata Generated and Maintained by ARC/INFO

The use of metadata in GIS is not new. ARC/INFO generates and maintains metadata on the spatial registration, projection, and tolerances of a coverage, grid, or TIN, and uses the metadata as vital input in the application environment. So, every time you start a project (i.e., create a workspace in ARC/INFO), the software automatically creates metadata, for example, an INFO directory and a LOG file. Similarly, every time you create a coverage, ARC/INFO automatically creates a set of files, for example, the TIC file (containing data about the coverage's coordinate registration), the LOG file (tracking all ARC operations performed on its respective coverage, grid, or TIN), the BND file (containing the coordinate values that denote the outer boundary or spatial extent of your coverage), and a TOL file (storing a coverage's processing tolerances). These metadata are also updated by the software with every operation that uses and modifies the respective values.

Additional ARC/INFO metadata relates to INFO tables. Some of the INFO table's properties can be found in the .DIR file in the INFO directory, while the descriptive details of the items within an INFO table are stored in the .NIT file. ESRI products manage only those metadata that are native to their data model. Any metadata stored in external relational database management system (RDBMS) (such as ORACLE, SYBASE, or INFORMIX) are currently not supported.

ARC Macro Language-Based Metadata Management

Another set of metadata is data dictionary-type metadata created and maintained by ARC Macro Language (AML™) based applications. A data dictionary (DD) is a complete listing of descriptions of all data items in a database, used to support an organization's operations. An example of such an application is the DD module of the ARC/INFO Application Development Framework (ADF™). These metadata describe the physical and logical view of the user's GIS database in

user language, independent of any application. For example, a tax assessor's database contains parcel boundaries, parcel IDs, assessed property value, and parcel owner, and represents them the way you see and understand the data in your database. The DD is populated by extracting metadata for the respective data items and stores them in a series of INFO tables. It includes descriptive or attribute data domains and domain value ranges. The DD module automatically updates this metainformation and supplies it to your application upon request.

Metadata, as described above, provide users with an informational view of their spatial database and their GIS application. In the following section, we describe how to extend metadata with additional, bibliographic-type metadata.

Metadata Capture with DOCUMENT.AML in ARC/INFO Version 7.0.3

ESRI has delivered an ATOOL called DOCUMENT.AML with ARC/INFO Version 7.0.3, to support the immediate metadata documenting mandate of federal agencies and needs of other organizations as well as open geodata access. The AML was jointly written by the USGS's Water Resources Division and the Environmental Protection Agency and has been adapted by ESRI to work with ARC/INFO. DOCUMENT.AML's sole purpose is to facilitate recording of bibliographic-type metadata.

Use DOCUMENT.AML to input, update, display, cross-reference, and output metadata for your coverages, grids, TINs, and library layers. DOCUMENT creates five metadata files: a .DOC file, a .REF file, a .NAR file, an .ATT file, and a file with site data. This last file contains source, ordering, and distribution information. Since its content frequently applies to more than one of your coverages, it is best stored and maintained at the directory level and cross-referenced to each coverage, grid, TIN, and library layer to which it applies. The .DOC file contains general information about your coverage, such as creation date, date of last update, number of arcs, theme name, and so on. Citation data for the source are stored in the .REF file. You can keep any descriptive information you deem noteworthy about your coverage in the .NAR file. For example, intended use, an abstract, as well as all information pertaining to the initial creation of the coverage.

The .ATT file describes object attribute data and individual data items stored in tables.

The user interface of DOCUMENT.AML is a series of form menus that provide you with a consistent and efficient frame for collecting metadata. Most metadata are entered by typing—a tedious and time-consuming, yet unavoidable, task. ESRI considers DOCUMENT.AML as an interim tool for your spatial metadata capture and storage. ESRI software development currently underway includes major improvements for metadata management on both the content and the functionality of DOCUMENT.AML. Details are provided below. At the same time, ESRI is improving the first version of DOCUMENT.AML for the next ARC/INFO release to better comply with the nomenclature, the content, and the output format of the FGDC Standard.

Future Metadata Management for ESRI Products

Traditionally, data dictionaries have been used to store nonspatial metadata managed in RDBMSs. Ways and means are now being explored on how the data dictionary concept can best be extended to the spatial domain to store and manage spatial metadata and how to use it in support of GIS applications (Seaborn, 1995). At this point in time, ESRI is not building on this concept; rather, we focus on management of spatial metadata within ARC/INFO and ArcView through a shared metadata management tool. To avoid any confusion with the traditional data dictionary concept, we have chosen metadata manager as the working title for the new functionality. Think of the metadata manager as a library that stores metadata management routines and that is shared by all ESRI products.

A note of caution is added here to let the reader know that the following section reflects in general terms the current status of product design. This design does not intend to suggest that the product will indeed be delivered as discussed here. Rather, as with any design, it is subject to changes necessary to provide our users with a high-quality software product.

Metadata Management Functionality

ESRI's first goal for adding the metadata management functionality to ARC/INFO and ArcView is twofold: first, to assist our users who have moved from building to using spatial data and need a tool to manage their spatial data and, second, to make ARC/INFO easier to use for users at all levels of expertise. Respectively, the metadata manager functions will handle two basic types of metadata. The first type is bibliographic metadata consisting of, and not limited to, the seven FGDC Standard categories listed earlier in this paper. ESRI aims to support an extensible, more flexible content than that of the FGDC Standard. The second type is ESRI metadata consisting of the specification of core objects such as a coverage, grid, and table in the ESRI data model. In addition, keywords, templates, and an easy-to-use interface for entering and retrieving metadata will be provided for both types. They will be stored in and be retrievable using the metadata manager, which is shared by all ESRI products.

Examples for keywords include abstract, publication date, status, bounding coordinates, theme, attribute accuracy, lineage and process step, map projection, distribution liability, native data set environment, and many more. An example of a template is the coverage template. The coverage template contains all elements that define a coverage and presents them in a consistent, standard form menu interface. You can customize the coverage template by specifying, for example, all the elements your road coverage needs, from its precision, feature types, and their attributes to the attribute domains (including nominal and ordinal values) and units of measurement, and others. Your roads coverage can in turn become the template for all road coverages created in your organization, which supports the use of standards discussed earlier in this white paper. These new product features aim to extend and improve the capability, content, and functionality of editing and mapping applications (as currently done by the ADF-DD) as well as recording bibliographic data (as currently supported by DOCUMENT.AML). The new metadata manager will also aim to support the concept of the framework data themes, which are a vital component of the NSDI's National Geospatial Data Clearinghouse (FGDC, 1995).

Adding New Data Types to the ESRI Data Model

Based on the preceding premises, the metadata manager extends the ESRI data model. The bibliographic metadata are in text file format of varying length. They will be stored in special documentation files and managed by ESRI products through an extension of the software using the shared metadata library. The ESRI metadata will focus on two aspects: defining existing core ESRI objects such as the coverage, grid, and TIN by way of their properties, relationships, and, ultimately, operations; and adding complete table item definitions including item domain specifications of nominal and ordinal type. In ArcView, metadata will be added as a new object class. Metadata management thus becomes an integral part of ESRI products and contributes to product unification.

By defining, storing, maintaining, and managing existing and new data types, the metadata manager presents the ESRI data model in a formal, coordinated, and consistent manner to our users. Thus, the software becomes easier to use and improves on some nagging issues in spatial data management. For example, access to these metadata is handled according to the content type. Metadata about a coverage or about a table specification are deemed critical to an application and therefore require fast and easy access. Bibliographic-type metadata, on the other hand, must be extensible in content to meet the varying requirements of our U.S. and international users.

Future Software Product Plans

Long-term products development seeks to further improve metadata management in two major areas. First, providing ready access to ESRI metadata benefits your use of ARC/INFO® software's core functions: editing, analysis, and mapping. Second, increased automation of standard, bibliographic metadata capture and update will reduce errors that result from manual metadata input and will help to increase spatial data set documentation, which, in turn, expands open geodata access, spatial data set sharing, and system interoperability.

In addition, a future version of the metadata manager aims at more data dictionary-type functionality. This includes support of full metadata documentation for all core ESRI objects during editing, analysis, and mapping operations. For example, you are editing a road coverage. Specification, storage, and access to feature editing

environment metadata—including all editing operations and constraints for lines of the specified road type and its subclasses—will allow you to perform your desired editing with the push of a button or selection from a scrolling list. Another example is cartographic object metadata such as the north arrow and scale bar, standard items in a map composition. Likewise, metadata for a variety of application scenarios, for example tax assessment, natural resource management, municipal government, and environmental monitoring, can be defined and stored for on-the-fly retrieval. A searchable access catalog, an integral part of such a data dictionary-type environment, is also envisioned for a future version of the metadata manager.

Other aspects in long-term product development may consider the possibility of, and requirements for, on-line metadata that document off-line spatial data or even not yet collected spatial data; the inclusion of a metadata image, browse image, or thumbnail as part of the bibliographic metadata to let you see for yourself what the available spatial data look like; and application domain-specific metadata, including database schemata, for example, for a local transportation system with vehicle routing or for an environmental monitoring system, and documentation of AML code.

These categories reflect the multidimensionality of metadata. More research and analyses are required to determine the interplay and role of these categories within the GIS environment. As ESRI products evolve and become more intelligent, spatial data handling becomes easier for you because spatial data management—including the automation of metadata capture—will increasingly become software-based.

Summary and Conclusion

The availability of reliable data is the most basic and critical factor in establishing a GIS. To see what is available requires that GIS users document their digital spatial data holdings and publish this bibliographic documentation. To ensure the widest possible dissemination and use of such spatial data documentation requires that it is composed using standard terminology, definitions, format, and procedures. These are the central concepts underlying metadata.

Likewise, defining, storing, and maintaining ESRI's core objects, their properties, relationships, and operations in our software, explicitly establish standards and make it ready to use right out of the box, while keeping it flexible and extensible to satisfy diverse user and application needs. Adding such metadata management functionality facilitates coordinated and consistent handling of the existing and newly added metainformation for all ESRI products. Thus, it improves overall spatial data management of ESRI's products and makes them easy to use for all users, regardless of their level of expertise. At the same time, the metadata manager strengthens the adoption of geodata standards and supports open geodata access, GIS interoperability, and the NSDI's Clearinghouse framework data themes. These are central goals for incorporating ESRI metadata and bibliographic metadata into the ESRI data model.

Suggested Readings

Please note that the references provided below are not intended to be exhaustive nor do they suggest any preferences or endorsement of content by ESRI.

- Anderson, J., 1991: Using a Data Dictionary Inventory Form: A Fundamental Tool in Documenting a GIS Database. In *Proceedings 11th ESRI User Conference,* 1, 179–192, Palm Springs, California.

- Bretherton, F., 1994: Reference Model for Metadata: A Strawman. In *IEEE Computer Society Technology Commission on MSS Metadata Workshop,* February, 1994, University of Texas, Austin.

- Chorafas, D.N., 1989: *Handbook of Database Management and Distributed Relational Databases,* Chapter 10: The Data Dictionary, 139–157, Blue Ridge Summit, Pennsylvania.

- Environmental Systems Research Institute, Inc. (ESRI), 1995: *Application Development Framework (ADF)—Training Workshop Manual,* Redlands, California.

- Environmental Systems Research Institute, Inc. (ESRI), 1995: The Spatial Data Transfer Standard in ARC/INFO, ESRI white paper series, Redlands, California (in preparation).

- Environmental Systems Research Institute, Inc. (ESRI), 1994: *GIS Approach to Digital Spatial Libraries,* ESRI white paper series, Redlands, California.

- Environmental Systems Research Institute, Inc. (ESRI), 1994: *Open Geodata Access—Through Standards,* ESRI white paper series, Redlands, California.

- Environmental Systems Research Institute, Inc. (ESRI), 1994: A Study on European Trends in Standards for Geodata Exchange, ESRI internal report, Redlands, California.

- Federal Geographic Data Committee (FGDC), 1995: *Development of a National Digital Geospatial Data Framework,* Federal Geographic Data Committee, Washington, D.C.

- Federal Geographic Data Committee (FGDC), 1994: *Content Standards for Digital Geospatial Metadata,* Federal Geographic Data Committee, Washington, D.C.

- Frew, J., L. Carver, C. Fischer, M. Goodchild, M. Larsgaard, T. Smith, and Q. Zheng, 1995: The Alexandria Rapid Prototype: building a digital library for spatial information. In *Proceedings 15th ESRI User Conference,* Palm Springs, California.

- Gittings, B.M., 1995: Internet's Role in the Collation of GIS Metadata. In *GIS World Sourcebook 1995,* 535–536.

- Griffiths, J-M., and K.K. Kertis, 1994: Intelligent, Self-Documenting Audit Mechanisms: An Extension of Metadata. In *IEEE Computer Society Technical Commission on MSS Workshop on Metadata for Scientific and Technical Data Management* (held May 15–18, 1994, at the National Archives, Washington, D.C.).

- Hansen, D., 1993: Extension of the Metadata Content Standard of the Spatial Transfer Standard to Other Information. In *Proceedings 13th ESRI User Conference,* 2, 83–93, Palm Springs, California.

- Jacqz, C., 1990: Using Index Coverages to Make Coverage Naming Transparent to the User. In *Proceedings 10th ESRI User Conference,* 1, Palm Springs, California.

- Johnson, D., P. Shelley, M. Taylor, and S. Callahan, 1991: The FINDAR Directory System: a Meta-model for Metadata. In Medykyj–Scott et al. (eds.) 1991: *Metadata in the Geosciences,* 123–137, Loughborough, Great Britain.

- Kuggeleijn, R., 1995: Managing Data about Data. In *GIS Europe,* 4, 3, 32–33.

- Lanter, D.P., 1990: Trimming Large Spatial Databases with Lineage Analysis. In *Proceedings 10th ESRI User Conference,* 2, Palm Springs, California.

- Lanter, D.P., 1990: *Lineage in GIS.* Technical Paper 90-6. National Center for Geographic Information and Analysis, UC Santa Barbara, California.

- Lanter, D.P., 1991: Design of a Lineage-Based Meta-Data Base for GIS. In *C&GIS*, 18, 4, 255–261.

- Lanter, D.P., 1993: A Lineage Meta-Database Approach toward Spatial Analytic Database Optimization. In *C&GIS*, 20, 2, 112–121.

- Lanter, D.P., and C. Surbey, 1994: Metadata Analysis of GIS Processing: A Case Study. In *Proceedings of the 6th SDH,* 1, 314–324, T.C. Waugh, R.G. Healey (eds.), 1994. Edinburgh, Great Britain.

- Lenz, E., and W. Mattingly, 1990: An Approach to Cataloging Digital Data. In *Proceedings 10th ESRI User Conference,* 2, Palm Springs, California.

- Lillywhite, J., 1991: Identifying Available Spatial Metadata: the Problem. In Medykyj–Scott et al. (eds.), 1991: *Metadata in the Geosciences,* 3–11, Loughborough, Great Britain.

- Marble, D. F., 1991: The Extended Data Dictionary: A Critical Element in Building Viable Spatial Databases. In *Proceedings 11th ESRI User Conference,* 1, 169–178, Palm Springs, California.

- Medykyj–Scott, D., I. Newman, C. Ruggles, and D. Walker, (eds.), 1991: *Metadata in the Geosciences,* Loughborough, Great Britain.

- Mikkonen, K., and A. Rainio, 1995: Towards a Societal GIS in Finland—ArcView Application Queries Data from Published Geographical Databases. In *Proceedings 15th ESRI User Conference,* Palm Springs, California.

- Miller, D., and B. Forner, 1994: Experience in Developing a Natural Resource Data Directory for New South Wales. In *Proceedings AURISA'94,* 391–398, Sydney, Australia.

- Miller, D., and K. Bullock, 1994: Metadata for Land and Geographic Information. An Australia Wide Framework. An Australia Standard. In *Proceedings AURISA'94,* 251–259, Sydney, Australia.

- National Academy of Sciences (ed.), 1995: *A Data Foundation for the National Spatial Data Infrastructure,* Washington, D.C.

- Nebert, D.D., 1993: Status of the National Geospatial Data Clearinghouse on the Internet. In *Proceedings 15th ESRI User Conference,* Palm Springs, California.

- Nebert, D.D., 1993: Implementation of Wide-area Information Server (WAIS) Software to Disseminate Spatial Data on the Internet. In *Proceedings 13th ESRI User Conference*, 2, 575–684, Palm Springs, California.

- Ostyn, F., 1995: The EDRA—Fueling GIS Applications with Required Geographical Information. In *Proceedings 15th ESRI User Conference*, Palm Springs, California.

- Seaborn, D., 1995: Database Management in GIS: Is Your System a Poor Relation? In *GIS Europe*, 4, 5, 34–38.

- Selle, A.R., 1995: Serving Spatial Data and Metadata from EPA Region VIII. In *Proceedings 15th ESRI User Conference*, Palm Springs, California.

- Tosta, N., 1994: Continuing Evolution of the National Spatial Data Infrastructure. In *GIS/LIS '94 Proceedings*, 769–777, Phoenix, Arizona.

- Trivedi, N., and T.R. Smith, 1991: *A Conceptual Framework for Integrated Metadata Management in Very Large Spatial Databases.* Technical Paper 91-2, National Center for Geographic Information and Analysis (NCGIA), University of California at Santa Barbara, California.

- U.S. Geological Survey (ed.), 1992: *Information Exchange Forum on Spatial Metadata.* Reston, VA (sponsored by Federal Geographic Data Committee, June 16–18, 1992).

- U.S. Geological Survey, 1992: *Spatial Data Transfer Standard (SDTS),* Reston, Virginia.

- Walker, D.R.F., 1991: Introduction to Metadata in the Geosciences. In Medykyj–Scott et al. (eds.), 1991: *Metadata in the Geosciences*, xiii–xvi, Loughborough, Great Britain.

■ Williams, D., 1995: Automated Capture of Metadata: Simple Procedures and Tools for Editing Coverages. In *Proceedings 15th ESRI User Conference*, Palm Springs, California.

ESRI GIS SOLUTIONS

For over 25 years ESRI has been helping people manage and analyze geographic information. ESRI offers a framework for implementing GIS in any organization with a seamless link from personal GIS on the desktop to enterprisewide GIS client/server and data management systems. Our GIS solutions are flexible and can be customized to meet the needs of our users. ESRI is a full-service GIS company, ready to help you begin, grow, and build success with GIS.

ArcView® *software enables users to quickly select and display different combinations of data and to creatively visualize information from their desktops.*

ARC/INFO® *software is a high-end GIS with capabilities for the automation, modification, management, analysis, and display of geographic information.*

PC ARC/INFO® *software is a full-featured GIS for DOS and Windows™-based PCs.*

ArcCAD® *software links ARC/INFO, the world's leading GIS software, to AutoCAD®, the world's leading CAD software.*

SPATIAL DATABASE ENGINE (SDE™) *software is a high-performance spatial database that employs a true client/server architecture to perform fast, efficient spatial operations and management of large, shared geographic sets.*

For more information, call
1-800-447-9778 (1-800-GIS-XPRT)
Send E-mail inquiries to info@esri.com
Visit ESRI on the World Wide Web at http://www.esri.com

CORPORATE OFFICE

ESRI
380 New York Street
Redlands, California
92373-8100 USA
Telephone: 909-793-2853
Fax: 909-793-5953

U.S. OFFICES

Alaska
Telephone: 907-344-6613

Boston
Telephone: 508-777-4543

California
Telephone: 909-793-2853, ext. 1906

Charlotte
Telephone: 704-541-9810

Denver
Telephone: 303-449-7779

Minneapolis
Telephone: 612-454-0600

Olympia
Telephone: 360-754-4727

Philadelphia
Telephone: 610-725-0901

San Antonio
Telephone: 210-340-5762

St. Louis
Telephone: 314-949-6620

Washington, D.C.
Telephone: 703-506-9515

INTERNATIONAL OFFICES

ESRI-Australia
Telephone: 61-9-242-1005

ESRI-Canada
Telephone: 416-441-6035

ESRI-France
Telephone: 33-1-450-78811

ESRI-Germany
Telephone: 49-8166-380

ESRI-Italia
Telephone: 39-6-406-961

ESRI-South Asia
Telephone: 65-735-8755

ESRI-Spain
Telephone: 34-1-559-4375

ESRI-Sweden
Telephone: 46-23-84094

ESRI-Thailand
Telephone: 66-2-678-0707

ESRI (UK)
Telephone: 44-1-923-210-450

For the location of an international distributor, call 909-793-2853, ext. 1375

Printed in USA