By Rj Zimmer, LS

Rj Zimmer is registered in Oregon and Montana, and has more than 25 years of surveying experience in the private and public sectors. He is the GIS Consultant & GIS Center Manager for the City of Helena – Lewis & Clark County Geographic Information Services Center in Montana.

Data Modeling of Survey Data in GIS

In order to store survey information electronically, data must be modeled in some way. A data model is a logical structure for the storage of information, and determines what information is stored, how the data are represented, and how each piece of information relates to other pieces of information within the data set or to other data sets. The concept of a data model should not be confused with the data format. The data *format* is the electronic storage form. Data formats include such electronic forms as ASCII, Dbase IV, XLS, etc. The data *model,* on the other hand, is how the pieces of information are structured, regardless of the data format. A good data model, in fact, is storage format independent, and can be ported, or used, in a variety of formats. For example, a data model for a survey point might be a simple flat structure, with only a few pieces of information per record, as shown in **Figure 1**. Typically, a text file of survey data uses one line of text for each record, where a record represents, for instance, a measurement or survey point. This data model can be formatted as an ASCII file, an Excel spreadsheet, a Microsoft Access table or any number of other formats. **Figure 2** shows this simple data model in two such formats, ASCII and Excel.

SIMPLE DATA MODEL:

- Point ID
- Description
- X-coordinate
- Y-coordinate
- Z-coordinate

Figure 1

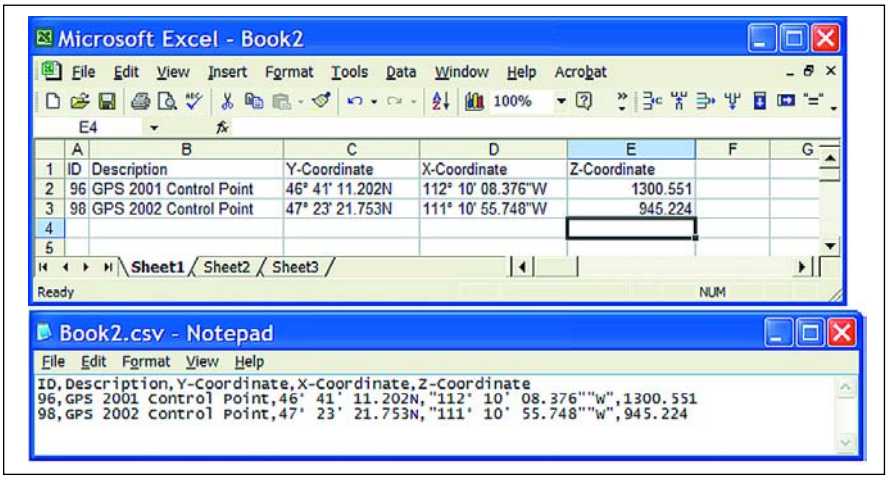

Figure 2 Format comparison for simple model

The difference between a data model and data format can be explained using the grade stake as an example. In grade staking there are certain kinds of information to convey and a medium used to convey that information. The wooden stake is the medium or format. The information that is written on the stakes is modeled in a particular way. The client will typically require a particular format, such as what data to put on the stake, how many stakes to use, where to locate the stakes, etc. For example, we may choose to model our construction information in a two-stake format: a grade stake and an offset stake (**Figure 3**). Each stake contains its own information and has a relationship to the other stake. How the information is posted to a stake is the data model.

Therefore, in general, the data model identifies the information that we wish to convey and how we chose to convey it. Data models can be simple or complex, depending on the nature of the information that is modeled and the intended uses for the information. In the example such as the survey point in Figure 1, the uses for the information may be limited due to the small amount of information contained in the data model. The use of the data might be to answer such simple questions as: How many survey points do I have and where are they? But as we shall see later on, survey data can be modeled in complex ways as well. The important considerations for data modeling are to understand the data itself, and how the data will be used. A typical data model may include a set of tables with different kinds of data in each table. The tables may be related through the use of keys (unique identifier codes) that are common in different tables. A multiple table format in a relational database structure can simplify the database and actually reduce the volume of data in the database. For example, a database might have a single list of survey points, and a single list of projects. The survey point, however, might be used in many projects. Instead of repeating all the informa-

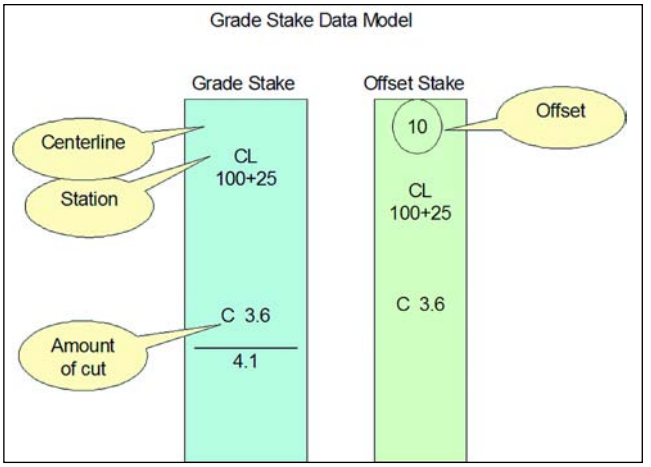

Figure 3 |

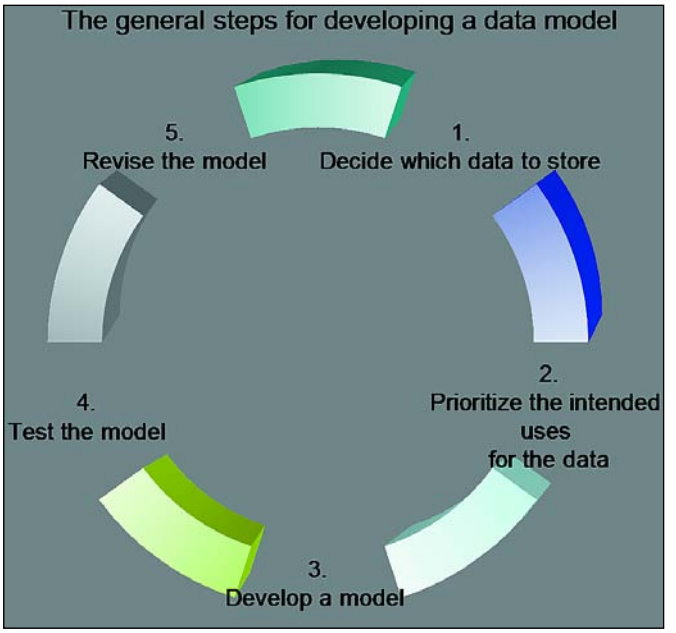

Figure 4

tion for each survey point within each project table, only the unique ID or key of each survey point needs to be listed. The application can then go to the survey point table to look up the other information about the point.

Modeling Survey Info

Although many data models exist for a variety of GIS applications and GIS data layers, the work on modeling survey data for GIS use has only begun for some types of survey data. **Figure 4** shows a typical process for developing a data model. Nevertheless, there is a variety of survey information that is used in geographic information systems. Some are fundamental survey data, such as measurements or control point information, and some are derived from survey data, such as land ownership or digital terrain models. Some of the common survey data used in GIS are:

- Cadastral Data
- Geodetic control point information
- Survey point information
- Survey measurement data
- Planimetric information
- As-builts of manmade features
- Terrain information (such as DTMs, contours, elevation points)
- Hydrographic (water elevations, flood elevations, bathymetry)

In addition to the fundamental survey information used in a GIS, there are also data used within survey information systems for internal uses by the survey firm, for things such as project management and marketing. These types of data might include information on projects (location, client, duration, cost, resources used, etc.)

Data models for survey and related information can be simple, as shown in the first example in Figure 1 for a survey point, or complex, such as the Cadastral Data Model or the NILS Project Survey Management System Mode (for more information on these projects see *www.blm.gov/nils/*). Although there may be other approaches to data modeling, generally the complexity of the model depends primarily on two things: the intended use of the data, and the nature of the data itself. One example of survey data GIS data model that is currently in use is the survey data model developed by ESRI for its Survey Analyst extension to ArcGIS.

The ESRI Survey Data Model

The survey data model used in the ESRI Survey Analyst extension is a complex one that models the survey information as a set of computations, coordinates, measurements, and survey points that are stored within the context of a survey project. One of the clever features of this data model is the ability to store multiple sets of measurements between–and derive (and store) multiple sets of coordinates for–survey points. This is a very adept model that allows the user to select which set of coordinates best suits a given application at any time. Moreover, due to the robustness of the data model, the survey data can be related to GIS features that are external to the database and thereby retain the feature's dependence on the fundamental survey information.

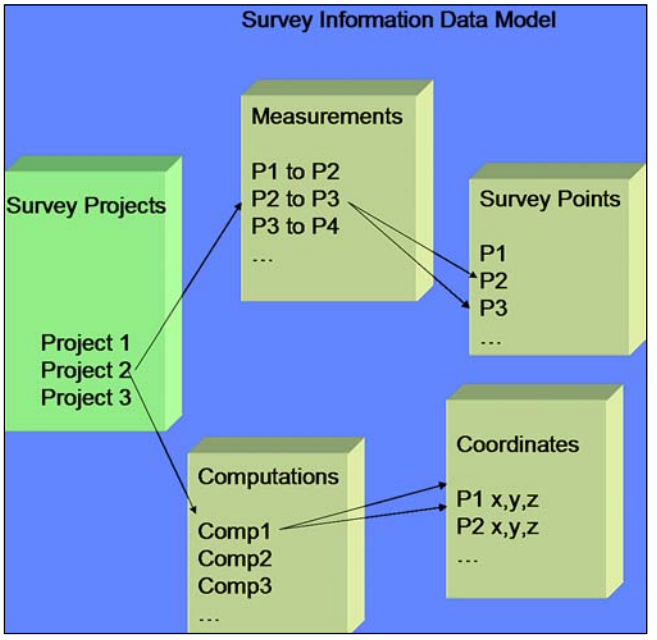

Figure 5

Whether simple or complex, it is important to remember that any desktop application that stores and/or uses survey data also models that data in some fashion. Some of these applications allow the export of the data to other formats, and typically do so to a choice of

continued on page 78

ZIMMER *continued from page 76*
formats, and perhaps even choice of models.

Distinctions

With digital data it is also important to understand the distinction between data models for the use or storage of information and data models for the exchange of information. How you store and use your data is up to you, however, when you want to get a copy of someone else's data or provide someone else a copy of your data, it may be necessary to use an intermediary or exchange format. A data exchange *standard* can be a way of providing uniformity in the content and types of data that are exchanged digitally. It defines the essential elements for giving and receiving data between parties. The exchange standard should support the most basic and fundamental use of the data and provide the receivers enough data and information to be able use it or understand enough about it to decide whether or not it will work for their purposes. The Federal Geographic Data Committee and the GeoSpatial One-Stop efforts strive to develop data exchange content standards for a variety of data, including cadastral and geodetic control data (go to *www.fgdc.gov/standards/standards.html* or *www.geo-one-stop.gov/* for more information). For example, the geodetic control content standard development is being led by the National Geodetic Survey, yet they are actually working on a standard that would accommodate control points other than the NGS Bluebook Datasheets. The goal is to provide a template for what information to provide someone else, regardless of how you store and use your data, or how they store or use the data that you provide.

Why a Standard Model?

As the desire by surveyors and non-surveyors to use survey data within GIS increases, data modeling of survey information becomes increasingly important. The ability to model the data properly so as to support the intended uses of survey data is important. It is also important to understand how a particular set of data is modeled in order to understand how to use the data properly. Standardized data models provide the benefits to facilitate the portability of data, simplify application development, make it easier to develop a data set, and increase the viability and flexibility of existing data.