Project Planning and Data Dictionary Design: Keys to Successful GPS Data Collection

James Stein, GIS Specialist, National Park Service Cultural Resources Geographic Information Services, 1201 Eye St., NW, Washington, DC 20005; james_stein@contractor.nps.gov

Abstract

The first step to successful GPS data collection in the field is project planning and data dictionary design. Completion of these components allows surveyors to conduct a focused survey that returns the needed data and reduces costly field time. This discussion will cover aspects of planning and executing a successful GPS project. Project planning addresses the use of GPS as an additional tool to efficiently collect data in the field in order to populate a GIS. Topics include assessing the purpose and data needs of the project, data scale and accuracy, feature attributes, and data dictionary development, as well as best survey practices.

Project planning

The first step, project planning and data dictionary design, allows surveyors to conduct a focused survey that returns the needed data and reduces costly field time. Also, the additional costs of post-processing and editing are minimized due to tailored data collection. The use of GPS as a tool to efficiently collect data in the field in order to populate a GIS is addressed in project planning. There are four steps that are key in planning a successful project; deciding on the purpose of the project, writing a project description, defining the level of accuracy needed, and determining data needs.

Projects can be broken down into three basic types; those meant to collect baseline data of resources, those designed as a specific application to answer questions about the resources, and projects that are a combination of the two. The baseline inventory project operates as an overview of resources. In this type of project, data reflects many types of features with more generalized and broad attributes about those features. Even though there may be no specific application for this type of database, it is often used as a base of context for more focused projects.

Projects designed for a specific application look at resources in a more focused way. These databases contain limited, targeted features with specific attributes. Often because of the lack of base data, a project must include components of an inventory-type project in order to be able to develop and apply the application project.

A project description explains the purpose of the project, any analysis to be completed, methodology to be used, data needs, deliverables, time and money budgeted, and the infrastruc-

Citation: Weber, Samantha, ed. 2012. *Rethinking Protected Areas in a Changing World: Proceedings of the 2011 George Wright Society Biennial Conference on Parks, Protected Areas, and Cultural Sites.* Hancock, Michigan: The George Wright Society.

ture, staff, software, and hardware needed to complete the project. Any analysis required should be spelled out in detail in the project description. The analysis and deliverables will directly inform the data needs and scale.

The purpose of the project and the type of analysis being completed are the primary variables influencing the scale of the data, but they are not the only factors. Usually baseline inventory data is collected at a scale that is comparable to existing base maps. These will vary according to organizational standards and the size of the area involved. Collecting highly accurate data to be used with small-scale base data, such as 1:25,000 maps, would not be cost effective. Data at this level of accuracy may very well be available from numerous other sources. Analysis that is being completed on a regional level requires data at a smaller scale. But if the project area is compact, such as a site plan, or consists of unique features not normally publicly available data, then a site visit for larger-scale GPS data collection may be the most appropriate mans of collecting needed data.

Once the data needs are determined and specified in the project description, and the required accuracy is established, then data availability can be investigated. Geographic information is available from many sources; digital databases may already exist for your study area. Prime sources for publically available data are the United States Geological Survey, U.S. Census Bureau and ESRI ArcGIS Online. Collecting GPS data is costly, both in time and money. Not only do you have field time to consider, but also the added burden of post-processing the data. Using GPS to populate the GIS database should be considered only if the data is not already available at the accuracy and specificity needed.

If it is determined that GPS data collection is necessary, then a project-specific data dictionary should be developed. A data dictionary is a list of features to map, and the attributes of those features that will be recorded. It serves as a guide to the surveyor in the field. Capturing too much data can result in complex maps which are hard to read and analyze, and can incur unnecessary expenses. Conversely collecting too little GPS data can lead to incomplete maps and faulty analysis.

Data dictionary design and development

There are numerous ways to build a data dictionary: you can build it in the GPS-specific software, such as Trimble Pathfinder Office; it can be replicated in a ESRI GeoDatabase structure; a GeoDatabase can be built in the GPS Analyst extension in ArcGIS and convert it into a data dictionary; you can use GPS Analyst to convert a data dictionary into a GeoDatabase; other software packages, such as CartoPac, etc., can be used and custom forms for phone and tablet applications can be utilized.

There are three steps in developing and building a data dictionary: identify the features to observed and mapped, determine the attributes to be recorded for each feature, and test the data dictionary. When looking at the features to map, the features should be divided into two categories, target features and reference features. The attributes collected for each will differ. Target features are those that are specific to the application, and the feature attributes collected need to be project specific. Reference features provide context and quality control, so the attributes can be more generalized. For example the location of artifacts collected in an archeological survey would be the target features. But roads, trails, and fence lines added to that map would help archeologists re-locate the artifacts, providing a context for the location of the artifacts. Reference features, such as hills and streams would allow analysis of the artifacts in a landscape context.

All features must be a line, point, or polygon. The decision of what option to select is dependent upon the intended use of the data, and the scale at which it will be displayed. All features have an area, but due to accuracy and scale issues, it is not always efficient to map polygons (the only option that has an area) for all features, when a point or line would do. It would be use-

less to record buildings as polygons for a small-scale, region-wide project, but if you are mapping for facilities management, using a point to represent a parking lot may not work out if the maintenance department needs to calculate the cost of repaving.

Once you have identified the features needed to meet the project objectives, you then need to specify which attributes are project oriented. You should only use attributes that are recognizable in the field, are observable and readily understood. You should limit target feature attributes to those that address the project, be sure attribute values are objective, avoid over-reliance on vague assessments (e.g., "good," "fair," "poor"), and make sure values are distinct from each other. Always have a unique identifier for all of the target features. And most importantly, when developing a data dictionary, make sure that the field surveyors are familiar with the resources being mapped. If you are recording architectural details on structures, do not send a herpetologist.

Attribute values can be recorded in numerous formats; menus, pick lists, radio buttons, image icons, numeric, text, and auto-generated formats like dates and times. In addition, certain functions, such as auto incrementing, can be programmed into the data dictionary. With formats such as menus, pick lists, radio buttons, you can standardize choices, reduce input errors, and make the data easier to query. But these formats are less flexible, although this can be mitigated by allowing multiple choices, and must be well thought out. If the surveyor runs into something unexpected in the field, then they are out of luck. On the other hand, text and numeric boxes are very flexible and responsive to the unexpected and unanticipated conditions. They can be hard to query due to spelling and case errors. Often the best data dictionaries are a combination of both of these formats, contain values such as 'other' or 'unknown,' and contain a text box for comments.

I look at a data dictionary as iterative. I often expect to revise them, by adding and subtracting features, attributes and attribute values, and refine them, by going from the extraneous to the essential, the subjective to the objective, and the vague to specific.

This is an era of shrinking budgets, manpower shortages, and ever tightening controls on travel expenditures. Following these simple guidelines for project planning and data dictionary design will help stretch project budgets by maximizing the efficiency of time spent in the field, and limiting the amount of data that needs post-processing and editing.