STANDARDS FOR TRANSMITTAL OF GIS DATA TO SHPO DATA SYSTEMS

D R A F T
Original draft: January, 10, 2002
Major revision: August 29, 2002

1/2002 Author-Collaborators
Laird Naylor, BLM Salt Lake City FO
Linda Clark, BLM Salmon FO
Kristen Jensen, Utah SHPO
Barbara Perry, Utah SHPO
Mary Hopkins, Wyoming SHPO
Eric Ingbar, Gnomon, Inc.
Mike Drews, Gnomon, Inc.
8/2002 Revisions
Eric Ingbar, Gnomon, Inc.
Mike Drews, Gnomon, Inc.
Mary Hopkins, Wyoming SHPO

Introduction

The use of GPS as a field tool and GIS as an office and management tool will continue to expand in the field of cultural resources management. Indeed, GIS itself will become more of a field tool in the next few years, with data available on handheld computers and in GPS data collection units. As more field observations are generated in digital formats, there is a natural desire to avoid re-digitizing data, including conveying spatial information directly into SHPO archives electronically.

GPS is revolutionizing how maps get made. Even low-cost receivers can calculate geographic coordinates more precisely than one can determine them from a 1:24,000 USGS topographic map. GPS information can usually be converted into a GIS file format compatible with SHPO spatial databases.

GPS field recordings, even those made with sophisticated data collectors, cannot usually be imported into SHPO data systems directly. The difficulty is not file formats, but in the nature of GPS data itself. GPS is a mapping tool, like a theodolite, an electronic distance meter, or a stadia rod. These tools require skillful manipulation of their results to create a coherent site map. So does a GPS. GPS does make collecting the information needed to make effective maps much easier.

SHPO's (and probably agency reviewers too) have little interest in receiving all of the GPS files from a fieldwork episode, just as they had no interest in receiving the instrument books from people mapping with alidades and plane tables. Cartographic representation was and is an interpretive activity. SHPO information systems need to receive cartographic representations of cultural resources phenomena in a format appropriate for error checking and rapid inclusion in their database and GIS archives.

The goal of this proposed set of standards proposal is to create a skeleton that can be used in almost every western U.S. state to convey records to a SHPO or agency GIS. The source of the records could be an agency field office, a contractor, or an avocational group. The standards discussed here pertain to conveying data as GIS datafiles. In general, we expect that organizations capable of creating GIS records are using GPS as a mapping tool. But, there is no necessary equivalence here: one could be using just GIS to create records. The standards proposed here are indifferent as to the source of spatial information; instead they force reasonable description of its accuracy.

Some terms used throughout this document are “phemonenon” and “entity”. Phenomonon is the real world “thing” that is mapped or recorded. Entity is the representation of a phenomonon in an information system. Usually, entity refers to a GIS entity (a graphic representation in a known coordiante space), but it could also be a row of data in a table.

GPS and the collection of spatial data

Most SHPO offices currently map site locations and boundaries to a 1:24,000 map base. While there are formal standards for map accuracy at 1:24,000 scale, the general notion in SHPO map files is that the cultural resources depicted are “pretty good” representations of site location and boundaries. This is inherent in the map scale. For example, a 0.5mm pencil line on a 1:24,000 map is 12m wide. A site represented as a visible 2mm dot would be 48m in diameter and cover an area of about half an acre (actually 1,806 square meters or 19,516 square feet).

GPS brings much higher accuracy to the recording process. Even uncorrected GPS from “sports model” GPS receivers will plot accurately at 1:24,000 scale. At larger scales (e.g., 1:6,000), inconsistencies will start becoming evident in uncorrected data.

Fieldwork gathers GPS and other spatial information in whatever way is most effective. Sites may be mapped using a variety of tools at the same time. The resulting raw spatial information In turn, the field mapping information is cleaned and made more regular to produce site “sketch” maps at scales of 1:100 to as small as 1:24,000 and site “location” maps at scales of 1:24,000 for inclusion in the site record packet.

For paper cartographic purposes, the draftsman can synthesize multiple sources of spatial information into a coherent single map. In digital cartography one is always tempted to retain the original digital data to the fullest possible extent, yielding a map that is complex and possibly difficult to interpret. For example, a site boundary could be created from GPS lines, GPS points, interpolations between points and lines, topographic lines from a USGS map, and a fenceline traced from a digital aerial photography. If one asked, “what is the spatial accuracy of the boundary?” there would be no simple answer.

The standards proposed in this document are use-oriented. That is, they aim toward the use of the information with which they are associated rather than documentation of its genesis. An earlier draft of this document requested for each spatial entity the types of GPS receivers, positional dilution of precision (PDOP) values, numbers of filtered positions, and other information about how each spatial entity was created. This revision of the earlier documentation moves description of field and office coordinate determination methods to a single metadata file for conveyed spatial data sets. Instead, it requests for each entity four attributes to facilitate use of the spatial data for most purposes.

The standards document

This standards document grew out of a working collaboration between several different organizations. Gnomon, Inc. staff served as the lead authors. In the absence of any adopting agency (at present), revisions and recommendations to the standard should be addressed to Gnomo, Inc., 1601 Fairview Drive, Suite F, Carson City, Nevada 89701, attn: Eric Ingbar (eingbar@gnomon.com).

First, this document describes metadata that should accompany each data transmittal. Second, standard columns for inclusion in each GIS dataset are described. Third, a brief discussion of data formats for transmittal are discussed. Fourth, the interaction between these standards and local, state, or regional conventions is considered. Fifth, examples of several common phenomenon – recording – reporting – entity scenarios and the resulting metadata and column values are given.

Metadata

Metadata is information about a dataset. Typically, it describes dataset derivation, methods, revision history, and content. The Federal Geographic Data Committee (FGDC) is the steward of federal agency metadata standards (www.fgdc.gov). A proposed content standard and metadata standard for cultural resources, sponsored by the FGDC, was published by the SHPO's of Wyoming, New Mexico, in collaboration with Gnomon, Inc. (colby.uwyo.edu/fgdcncptt.html).

In general one is always well-served by following the FGDC standard, the Content Standard for Digital Geospatial Metadata (CSDGM) in release 2.0 as of this writing. Several tools are available to assist in the creation of formal metadata documents that meet this standard. An extension for ESRI's ArcView 3.x is available from the National Oceanographic and Atmostpheric Administration (NOAA). ESRI's ArcCatalog/ArcGIS contains a very flexible metadata tool. Each of these tools allows one to create and store boilerplate metadata entries, so that one can generate similar documents swiftly once initial values are entered. Numerous other tools are tracked at the FGDC web site.

The FGDC standard can appear onerous to the casual data submitter. Whether one uses this standard or not, there are some items that should accompany every data submission. These are shown in the following table. Metadata documents must be sent in plain text formats. Optionally, one might also include the same information in formats created by metadata tools.

Topic	***Description***
Data creator	Company, agency, or other organization who created the data in the dataset.
Date created	Date on which the dataset was created, finalized for conveyance
Associated activity, resource identifiers	A list of identifying numbers associated with this dataset. Typically, this might be an organization project number, an agency investigation number, and a SHPO activity or project number. The purpose is to lead the user to appropriate paper records.
Methods, data processing description	Methods and data processing techniques used to create the data. A brief description will suffice This topic can include the field procedures, equipment, and protocols used for collecting spatial data. For GPS data collection, this could include receiver make and model, PDOP cutoff, etc. Post-field processing, aggregation, digitization, and smoothing may be described in this section.
Responsible party and point of contact in creating organization	The name(s) and contact information of a person in the data-creating organization who is familiar with the data and responsible for its quality.
Coordinate system, units, and datum of data	This topic must cover the coordinate system and datum in which the data are conveyed (not necessarily the coordinate system in which the data was created – this might be covered in the “Methods” section of the metadata document, including conversion from the source coordinate system to the conveyed coordinate system). Different SHPO's may have specific requirements for data that will be accepted.

Proposed standard columns in GIS attribute tables

The following sections describe mandatory columns that shall be present in each GIS attribute table and must be populated with attribute values for each row in the table. We have not included a full description of entity labels in the mandatory columns, since these are variable from one state to the next. For each column attribute, except entity labels, a table of allowed values and what they denote follows the attribute description. For ease of reference, each column is presented on its own page.

Horizontal Positional Accuracy

Column name: **HposAcc (character, 6)**

Description: This attribute describes the *horizontal* positional accuracy of the GIS entity. Accuracy can be conceptualized as the likelihood that a stated coordinate is the true coordinate of a position. Hence, accuracy is the converse of positional error. The values for this attribute are probable positional error circles – the root mean square (RMS) error of a position.

For a single position, the root mean square error is a clear measure of accuracy probability. Many GPS units and post-processing software return RMS errors for averaged position fixes. However, there are many cases in which RMS error is more difficult to determine. RMS error is an estimate derived from repeated measures; single position fixes must default to having an RMS error at least as large as the usual RMS error for the source of the position fixes.

A practical guideline for determining the value of this attribute is the source of the coordinates ("location") used to create the GIS entity. So, a map coordinate measured from a USGS 1:24,000 quadrangle has an RMS error greater than the paper map itself.
Recommendations are provided below and in the examples.

Combining data with different horizontal positional accuracy can yield unexpected results. For example, a highly accurate GPS-determined position may plot on the "wrong" side of a USGS digital map image registered in coordinate space. This is not due to an error in either position, just the positional error in the map is greater than the error in the GPS position. Adjusting the GPS-position to match the map would degrade the horizontal positional accuracy of the GPS reading (and the HposAcc value should be changed accordingly).

Attribute values: Attribute values are the roughly 90% probability that a stated coordinate lies within a certain distance of the true coordinate.

Value	Example methods used to determine coordinate(s)
<1m	Averaged, differentially corrected high-end resource grade GPS; Survey-grade GPS; Experienced operator using 10" or more precise total station or theodolite and EDM traversing from a known coordinate monument less than 5000m distant
<10m	Single position of high-end resource grade GPS; multi-position averaging of sports-grade GPS without differential correction
<20m	Typical sportsman grade GPS – single position fix; USGS 1:24,000 map (National Map Accuracy Standard is approximately 13m)
<100m	USGS 1:36,000 to USGS 1:125,000 map
UnkLow	Unknown – low confidence in horizontal positional accuracy; likely error is not known, location is only an estimate quite likely to be erroneous
UnkHi	Unknown – high confidence in horizontal positional accuracy; likely error is not known, but coordinates are likely to be correct on a 1:24,000 scale map
UnkUnk	Likely error is not known and no estimate of reliability of horizontal position is possible

Horizontal Positional Source

Column name: **HposSrc (character, 10)**

Description: This attribute describes the source of the coordinates used to place the GIS entity into coordinate space. The attribute values describe only the most common sources and are not intended to be comprehensive. Horizontal positional source is useful as a means to segregate GIS entities derived from different sources, especially in data derived from plots on paper maps.

Attribute values: Attribute value is determined by the *source* of the horizontal coordinates. A GIS entity may have multiple sources, in which case one should state the predominant source. Multiple source entities that have no dominant source should receive an attribute value of “other”.

Value	Example methods used to determine coordinate(s)
GPS	A GPS unit, of any grade, was used
SurvInst	A total station or a survey instrument (transit, alidade, theodolite, electronic distance meter, stadia rod, or chain/tape), was used
USGS map, scale, e.g., USGS24000 USGS62500 USGS100000	Horizontal position coordinates were derived from USGS map at given scale. Note that if one transfers a GPS position to a map, then digitizes from the map, the accuracy is still that of the map, not the GPS.
Aliquot	Derived from an aliquot (cadastral) location. This depends upon the size of the aliquot part relative to the entity coordinate. At best, since an aliquot must be mapped to be converted to coordinates, the horizontal positional accuracy is that of the associated map.
Asserted	Horizontal position is an assertion with no other source information (e.g., a site record). In this case, horizontal positional accuracy will probably be unknown.
Other	Some other source, known but not among choices above.
Unknown	Source is not known.

Boundary Precision

Column Name: **BndPrec (character, 6)**

Description: Boundary Precision is the “fuzziness” or uncertainty of a reported boundary. It applies only to polygonal (having the geometric property of area) GIS entities. “Fuzziness” can be thought of as how sharply a bounding line should be drawn. An inaccurate boundary would be represented as a wide gray line, a very accurate boundary as a thin, darker, line. Current GIS display technology does not do a particularly good job of displaying the uncertainties in data, containing no display utilities by which one can indicate uncertainty or fuzziness easily. Note that the concept of boundary precision does not, generally, apply to whether a boundary is real, imagined, or how it was estimated. Rather, Boundary Precision is the reliability one places upon the boundary as a set of coordinates. Some other means is necessary to determine whether one should trust the method by which the boundary was defined on the ground.

Boundary precision can be a complicated estimation if one considers all of the potential error sources and uncertainties that compose a bounding line. For example, if one creates a boundary by joining together high accuracy GPS positions, what is the “fuzziness” of the lines between the points? This will depend upon how closely the position fixes fit the intended boundary. Too few points, or points in the wrong place, and a boundary can be quite inaccurate.

Many GIS polygons are composed of heterogeneous boundary sources, each of which could have its own spatial inaccuracy. In the standard presented here, each GIS polygon is given a single value for boundary precision. A more complex standard would involve associating appropriate accuracy attributes with *each* part of a polygon boundary. Although perhaps desirable, individual boundary segment attributes would be complicated to create and manage. For this reason, they are not incorporated into this standard. A single estimation is requested in this standard.

In general, the predominant technique used to gather or create coordinates of the ***observed*** boundary of an entity determines the boundary precision. Estimated parts of boundaries are not to be included in the estimated accuracy. Because of its inherent complexity, boundary precision will always be a judgment of the cartographer creating the GIS entity. The precision of the boundary

Attribute values: Attribute values are the estimated, appropriate values for a “gray” line to represent the boundary of a phenomenon, were a GIS to draw the boundary as a zone of probability. The values are intended to be best judgement, realizing that one will probably be combining different error widths in most cases.

Value	Example methods used to determine boundary
<1m	Averaged, differentially corrected high-end resource grade GPS; Survey-grade GPS; Experienced operator using 10” or more precise total station or theodolite and EDM traversing from a known coordinate monument less than 5000m distant
<10m	Single position of high-end resource grade GPS; multi-position averaging of sports-grade GPS without differential correction
<20m	Typical sportsman grade GPS – single position fix; USGS 1:24,000 map (National Map Accuracy Standard is approximately 13m)
<100m	USGS 1:36,000 to USGS 1:125,000 map
UnkLow	Unknown – low confidence in horizontal positional accuracy; likely error is not known, location is only an estimate quite likely to be erroneous
UnkHi	Unknown – high confidence in horizontal positional accuracy; likely error is not

	known, but coordinates are likely to be correct on a 1:24,000 scale map
UnkUnk	Likely error is not known and no estimate of the accuracy of horizontal position is possible

Boundary Observation Completeness

Column name: **BndComp (character, 8)**

Description: The Boundary Observation Completeness Attribute describes whether the boundary in the shown in the data represents the entirety of the boundary of the entity being mapped or only part of the entity boundary. The attribute is particularly useful in situations where only part of a phenomenon (e.g., a resource, an investigation) is mapped in the field. The attribute flags the observational completeness of the phenomenon boundary representation, not the logical completeness of the boundary. A boundary is logically complete simply by closure (for a polygonal entity); observational completeness means that the logical boundary matches the actual boundary.

Note that in the case of a linear or point phenomenon, an observed boundary may take the form of a line (perhaps the centerline of the phenomenon) or a point (perhaps the central point in the phenomenon).

A cultural resources example may clarify the concept of observational completeness. Consider an archaeological site recorded within a highway right of way. The crew recording the site is not allowed to leave the right of way, although the site runs outside of the right of way. So, the crew maps the boundary of the site as they observe it right up to the edge of the right of way. In the GIS, the polygonal shape representing this archaeological site is "squared off" at the right of way edge – the GIS entity is logically complete. In other words, there is a boundary represented in the GIS, but the entire boundary was not observed, so it does not represent the boundary of the entire phenomenon.

In practical terms, someone using the spatial data is given a means to determine whether the data are complete for a given phenomenon or whether the boundary shown is closed merely by convention. This determination is often very important for "linear" entities, such as roads, trails, or ditches. Segments of these phenomena may be recorded in their entirety (see below under the Segment attribute), even though the entire road, or ditch, or pipeline is not completely observed. Yet, each segment recording is a complete observation. No part of the reported segment boundary is an inference.

Attribute values: Attribute values for Boundary Observation Completeness signal to the user whether a boundary was completely observed or not.

Value	Boundary observation completeness
Complete	Entire phenomenal boundary was observed (mapped) completely. Note that this could mean the centerline of a phenomenon, or a centerpoint. By convention, a distinctly identified segment of a phenomenon can have a value of "Complete" if it has been mapped entirely.
Partial	Only part of the phenomenon was mapped.
None	The phenomenon boundary was not observed or mapped at all. Boundaries created by buffering using a convention (e.g., "sites less than 30m in extent shall be mapped as point and buffered to be a polygon 30m in diameter") would have a value of "None" for Boundary Observation Completeness.
Unknown	The observational completeness of the phenomenon spatial data is not known.

Segment

Column name: **Segment (character, 1)**

Description: The segment variable describes whether the spatial entity represents a definable segment, lobe, or part of an entity, rather than the entire entity. It is necessary because some phenomena are so extensive spatially, forbidding of access for mapping and observation, or otherwise unobservable. A spatal data user may be presented with a spatial entity of high positional and boundary accuracy, complete boundedness, and yet be shown only part of the entire phenomenon. This attribute flags such a condition for the spatial data user. If true (the entity is a segment), then there is more of the same phenomenon, perhaps present in data as different spatial entities. If false, then the entity represents the entire phenomonon.

The State of Hawaii is an excellent example of the use of Segment. In a GIS, the state is represented as several distinct polygons. Each is a part (a segment) of the state but is not the entire state. Thus, Segment = TRUE for each island.

Linear phenomena are particularly amenable to segmentation in spatial datasets. Simple examples abound: the portion of a highway that lies within a particular county. A railroad construction shoo-fly.

A lobe of an entity is also a segment. Above, the example of an archaeological site in a highway right of way was given (see Boundary Observational Completeness). The portion of the site within the right of way is a lobe, or segment, of the entire site. So, not only was the boundary incompletely observed, but the spatial data represents a segment of the entire phenomonon.

There is no necessary relationship between Boundary Observation Completeness and Segment. A boundary may be complete, but the entity is only a segment (e.g., a single Hawaiian island). A boundary may be incomplete and the entity is only a segment (the highway right of way site example). A boundary may be complete and the entity is not a segment (the entity represents the entire phenomon). A boundary may be incomplete and the entity is not a segment (for example, a partly observed archaeological site bound).

Attribute values: Attribute values for Segment are straightforward.

Value	Segment
T-TRUE	The entity is a segment of the entire phenomenon.
F-FALSE	The entity is the entire phenomenon.
U-UNKNOWN	The relationship of the entity to the phenomonon is unknown.

Maximum Entity Width applies only to point and linear entities

Column name: **EntWidM** (long integer)

Description: In many cases, phenomena are recorded in GIS as points and lines, even though they are two-dimensional (i.e., they have area). The Maximum Entity Width column gives a single metric value representing the width of th e entity. If one were to create a spatial boundary around the entity, then one would use half the Maximum Entity Width as a buffer distance.

When a phenomenon is presented in GIS as a polygon, it will not have an entity width (the value should be zero).

Attributes: The values in this column are the actual width or diameter of the phenomenon's spatial extent.

Value	***Description***
0	Entity should not be buffered to create a polygon (e.g., it already is a polygon)
1 to any value	Width or diameter of resulting polygon if entity is buffered to create phenomonon of appropriate size. Buffer distance in most GIS software would typically be half of this value.

Recommended file formats

Because polygonal shapes can overlap and create problems in topological GIS formats, a non-topological format is the most appropriate way to convey spatial data in this standard. There are several formats available, but the most widely used at present in federal and state agencies is the ESRI Arcview 3.x shapefile format. This format is non-topological and available as an export format choice in many software applications in addition to those from ESRI. Trimble Pathfinder Office, AutoDesk AutoCad Map, MapInfo, Intergraph GeoMedia, and many other applications can create this format.

The current recommended format is the ESRI ArcView shapefile format version 3.x, due to its widespread use. Ultimately, a neutral format such as GML-XML that is widely avalable (which GML-XML is not) would be the preferred form.

The role of local conventions

Regional, state, and local convention may override some of the standards above. When this occurs, it is incumbent upon the overriding organization to document how the standards described above should be met. For example, a particular state may require that all archaeological sites less than an acre be conveyed to them as points, not polygon GIS entities. So, even if one used a very accurate GPS to create a boundary for a less than one acre site, one must still determine a rough centerpoint and report its coordinate (as a point entity). Metadata and standard column entries for this should be provided by the *mandating* organization, rather than relying upon the information providers to document this convention adequately or consistently.

Another local convention might be the preference for different terms in the attribute values. We would rather see a table of equivalence created for this purpose, instead of changing the values shown here. So, a local term for positional accuracy might be “LOW” as opposed to the standard term “UnkLow”. We recommend that information be conveyed in the standard format and converted as needed by the receipient.