DATA STANDARDS FOR

MAINE GEOGRAPHIC INFORMATION SYSTEMS

Presented to the Information Services Policy Board by
GIS Executive Committee Chair, Richard Dressler
June 27, 2002

Approved by the ISPB
June 27, 2002
TABLE OF CONTENTS

1. INTRODUCTION

2. MAP COMPILATION/RECOMPILATION
   2.1 Map Scale
   2.2 Map Media
   2.3 Coordinate Reference
   2.4 Projection
   2.5 Line Work
   2.6 Coincident Features
   2.7 Data Not Recompiled on Standard Basemaps

3. DATA AUTOMATION SPECIFICATIONS
   3.1 Automation Technique
   3.2 Compatibility
   3.3 Digital Tolerances
   3.4 Cartographic Accuracy
   3.5 Data Capture
   3.6 Spatial Topology
   3.7 Edge Matching
   3.8 Coding Accuracy
   3.9 Accuracy Assessment
   3.10 Proof Plots
   3.11 Attribute Coding Schemes
   3.12 Database Development

4. DOCUMENTATION
   4.1 Basics of FGDC
   4.2 Technical Support for FGDC
   4.3 More About FGDC

5. REFERENCES
1. INTRODUCTION

This document provides guidelines covering the digital conversion of geospatial data into geographic information system (GIS) format. The overall goal is to ensure that a high quality, well-documented GIS database is built for the State of Maine. The document objectives are three-fold:

1. Set technical specifications for geospatial data automation and development;

2. Provide basic guidance in map compilation/recompilation; and

3. Provide standard procedures for documenting the history of each geospatial data layer and source map to aid users of the Maine GIS database in determining the viability of those data for specific applications.

All proposals, agreements, contracts and grants for GIS data automation should require that these standards be met or exceeded.

If agencies are planning projects that will generate large amounts of digital data, it is recommended that a pilot project be performed on a small area covering all aspects of the process from map compilation to the actual application. A pilot project insures that all systems and procedures used do in fact generate adequate data. The project should cover at least two study area modules (e.g. quad sheets) so that edge matching capabilities can be assessed. GIS personnel should participate in all aspects of the pilot project.

2. MAP COMPILATION/RECOMPILATION

The ultimate quality of a GIS database is as dependent on the quality of the source map as it is on the care used in the digital automation process. Whether new map data are being compiled onto a basemap or old maps are being recompiled onto a new base, certain basics need to be addressed. Precise automation procedures, although very important, will not improve an inferior source map. Georeferenced, rectified photographs and images, as well as vector basemaps may serve as source maps. Important considerations are:

2.1 Map Scale

- Varies with intended use and companion data layers
- Data submitted for inclusion in the Maine GIS database will conform to National Map Accuracy Standards (NMAS) for scale.

The choice of a basemap is the first step in the map compilation/recompilation process. Mapped data at any scale may be input to the system, but digital maps retain the accuracy of their original source map regardless of the scale at which they are plotted. Thus, a 1:100,000 map retains its original accuracy even when plotted at a 1:24,000 scale. The map scale also determines the best use of the final data. The following table illustrates how applications vary depending on the scale of the original map.

<table>
<thead>
<tr>
<th>MAP SCALE</th>
<th>MAP UTILITY</th>
<th># MAP SHEETS TO COVER MAINE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small scale 1:500,000</td>
<td>Statewide studies, planning</td>
<td>1</td>
</tr>
<tr>
<td>1:250,000</td>
<td>Statewide studies, planning</td>
<td>13</td>
</tr>
<tr>
<td>1:100,000</td>
<td>Regional studies, planning</td>
<td>35</td>
</tr>
</tbody>
</table>
USGS 7.5 minute quadrangle series have served as the standard digital basemap for agency data
development since 1990. The 1:24,000 scale properly balances the economic and accuracy
considerations of data development for most planning and natural resource activities. Basemaps of larger
scales or different sources may be considered, but the decision must be carefully judged relative to the
intended use, the scale of currently available digital data and the proposed basemap’s quality.

Data submitted for inclusion in the Maine GIS database will conform to National Map Accuracy Standards
(NMAS). For maps on publication scales larger than 1:20,000, not more than 10% of the points tested will
be in error by more than 1/30 inch, measured on the publication scale; for maps on publication scales of
1:20,000 or smaller, 1/50 inch. The following table shows the NMAS for some common map the scales:

<table>
<thead>
<tr>
<th>Scale</th>
<th>Engineering Scale</th>
<th>NMAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:500,000</td>
<td>1&quot; = 7.89 miles</td>
<td>+/- 833.33</td>
</tr>
<tr>
<td>1:250,000</td>
<td>+/- 416.66</td>
<td>appx. 127</td>
</tr>
<tr>
<td>1:100,000</td>
<td>+/- 166.67</td>
<td>appx. 51</td>
</tr>
<tr>
<td>1:63,360</td>
<td>1&quot; = one mile</td>
<td>+/- 105.60</td>
</tr>
<tr>
<td>1:24,000</td>
<td>1&quot; = 2000’</td>
<td>+/- 40.00</td>
</tr>
<tr>
<td>1:12,000</td>
<td>1&quot; = 1000’</td>
<td>+/- 33.33</td>
</tr>
<tr>
<td>1:4,800</td>
<td>1&quot; = 400’</td>
<td>+/- 13.33</td>
</tr>
<tr>
<td>1:1,200</td>
<td>1&quot; = 100’</td>
<td>+/- 3.33</td>
</tr>
</tbody>
</table>

A map resulting from the combination of two or more maps, photographs, or images of differing scale
retains the accuracy of the least accurate source map. It is, therefore, important to consider the scales of
data layers you expect will be most commonly used with the maps being prepared when choosing a scale
for your data. This will have a bearing on both the technical and economic aspects of a project.

2.2 Map Media

- Stable-base mylar

Whenever possible, the mapped data should be drafted on the most scale-stable medium available.
Shrinking and swelling in non-stable base media due to changes in atmospheric conditions can have a
profound effect on the spatial accuracy of the final digital product. The following represents, from most to
least desirable, the media to be used as the mapped data source for automation:

1. Mylar original
2. Mylar contact reproduction from mylar original
3. Non-stable base paper from mylar original
4. Non-stable base paper

Manuscript maps should not be folded. They should be stored in flat files or rolled up in map tubes.

2.3 Coordinate Reference
o Minimum of four reference points

Each map, photograph or image included in the Maine GIS database through digitizing or other means must have a minimum of four reference points (tics) for which the geographic coordinates are known and are printed on the map. Coordinates may be latitude/longitude pairs, or other units consistent with the projection of the map. If no coordinate reference is provided on the map, reference points must be calculated. Increasing the number of “quality” reference points will increase the accuracy of the registration.

2.4 Projection

o Projection of a map must be known and defined.

Prior to digitizing, and for later digital manipulation of map information, the projection of a map must be known. Maps are printed in many different projection systems. For example, most U.S. Geological Survey maps at 1:24,000 and 1:62,500 scales are in the Transverse Mercator or polyconic projection, although many other projection systems can be handled by current GIS technologies. Once in digital form, a map can be transformed from one projection to any other. The current standard projection for data in the Maine GIS database is Universal Tranverse Mercator (UTM), North American Datum 1983 (NAD83), Zone 19, Meters. To facilitate inclusion in the Maine GIS database, data should be submitted in this standard projection and projection should be defined.

2.5 Line Work

o Maximum line width of 0.013 inch

Lines to be digitized will be drafted with a standard 00 technical drafting pen producing a line width of no more than 0.013 inch. Lines that are too thick introduce subjectivity on the part of the digitizer operator.

2.6 Coincident Features

o Compile/recompile coincident features to avoid topological error.

Coincident features are those that are common to two or more data layers. For example, the mean high water mark of a pond may serve as a boundary for a residential zone, a soil type, and a wildlife management area; the mean high water mark, soil boundary and management area boundary are coincident features; therefore, coincident features should be digitized only once. Regardless of the care taken in digitizing, slight differences may lead to problems with small “slivers” if the layers are topologically joined. To avoid this phenomenon, use and reuse coincident features in new data layer development.

Coincident features should be clearly symbolized and labeled. During manual map compilation or reccompilation a coincident feature should be symbolized in drafting and labeled by source. A standard method is to draft coincident features with a dashed line. If a coincident feature is to be extracted from an existing digital map and added to a new digital data layer, it should be attributed with information on the source from which it is extracted. Dashed line symbology may also be incorporated as an attribute value to allow the feature to be displayed and/or selected as a coincident feature.

2.7 Data not Recompiled on Standard Basemaps

In some cases it may not be feasible to reccompile existing data on a standard basemap. The preparation of these data for entry into the Maine GIS database must be considered on a case-by-case basis with GIS personnel. The issues of media, coordinate reference, projection, and line work are particularly important under those circumstances.
3. DATA AUTOMATION SPECIFICATIONS

3.1 Automation Technique

- Features automated by heads up digitizing will be attributed with information on the source, and the process, used to add the feature to the data set and will equal or exceed NMAS standards for scale, of the original map, photograph, or image.

Tablet digitizing and scanning are currently the most commonly used techniques for automating maps. The selection of an automation technique depends on issues such as the complexity and quality of the source map, type of scanner, and the number of maps.

Data automation by heads-up digitizing has become more common with the increased availability of online scanned high resolution imagery. This is an interactive process in which data is created using previously scanned and georeferenced images such as scanned topographic sheets or photography. The process is similar to conventional tablet digitizing, but rather than using a digitizing tablet and a cursor, the user creates the data layer on the screen with the mouse and typically with referenced imagery as a background. The attention of the user is focused up on the screen, and not on a digitizing tablet. The accuracy of heads-up digitizing as a method of data automation depends on the type of image being used and the experience of the operator. For example, drafting features from a scanned topo sheet is easier to do with precision because the features have already been cartographically interpreted and drawn. By contrast, accurate drafting of features from a photograph requires the use of more photo-interpretation skill.

3.2 Compatibility

- Final format will be compatible with this standard.

As Maine’s state government GIS network evolves, emerging systems will use ARC/INFO or will be fully compatible with ARC/INFO through a common data interchange format so existing digital spatial data can be shared.

Therefore, all data developed for the Maine GIS database must be in a format that is easily converted to an interchange format compatible with other components of the statewide GIS.

3.3 Digitizer Tolerances

- Strictly follow software and hardware vendor guidelines for digitizing

There are several digital tolerances which affect the accuracy and resolution of a digital map that can be explicitly defined during the map automation process. The tolerances recommended by the vendor should be strictly adhered to when digitizing or editing data on a particular system.

- RMS error not to exceed 0.005

The Root Mean Square Tolerance (RMS) defines the error incurred when predefined tics are used to register a map on the digitizing board for automation. In order to maintain the spatial relationship of map features during digitizing, the RMS error must be kept as low as possible when a map is registered on the digitizing table. The recommended maximum RMS error is 0.005.

- Tolerance equivalent to 0.005 digitizer units

The tolerance defines the minimum distance separating line coordinates, and the tolerance set corresponds to the resolution of the digitizer.
3.4 Cartographic Accuracy

- Digitize map features to within the equivalent of 0.01 inches of the original manuscript.

Digitizing spatial data involves taking a hard copy map and tracing it with a cursor to produce a digital file. Errors and distortions can easily occur in the tracing process. These errors are often small or negligible, but this may not always be the case. Careful, consistent and systematic digitizing, plus thorough verification are essential to adequately retain the quality of the source maps.

- Scanned data will be georeferenced and vectorized to within the equivalent of 0.01 inch of the original manuscript

- Data collected with Global Positioning System (GPS) equipment will have an accuracy equal to or greater than the base data to which it will be added.

Base stations and differential correction should be used to maximize the accuracy of GPS data collection. An accuracy assessment and documentation of method, collection parameters, and technology that affect data accuracy should be provided with the data.

Required standards for feature accuracy are:

- 90 percent of the planimetric features on the digital map will be within 0.01 inch of the centerline of that feature on the manuscript map when plotted at the original scale.

- 100 percent of all features will be within 0.02 inch.

The 0.01 inch interval is equivalent to a standard 0.01 plotter pen width. When a proof hardcopy plot of the digital map is overlaid on the original basemap, discrepancies in line work will be seen as an open space between the plotted feature and the original manuscript. Discrepancies in point data are more difficult to judge.

It should not be assumed in conversion between data formats, i.e. shapefile to coverage or coverage to SDE layer, that cartographic accuracy will be preserved. Careful management of tolerances and precision of the work environment will greatly improve the results of such a conversion and the conversion process should be documented.

3.5 Data Capture

- Use a minimum number of coordinates to define a line.

Storing many layers of statewide digital data requires efficient use of available computer storage, so digitize the minimum number of vertices needed to accurately represent the cartographic feature within the 0.01 inch accuracy limit.

- Digitize the exact center of a point map symbol

If possible, point locations should be entered directly into the system for exact known coordinates such as latitude and longitude rather than by digitizing. When digitizing points from a source map, digitize the exact center of the point map symbol.
3.6 Spatial Topology

- No overshoots
- No slivers
- No open polygons (e.g. undershoots)
- No label errors
- No unresolved node errors
- No unresolved line segment intersections

Digital data submitted for inclusion in the Maine GIS database will be topologically clean and free of errors. All points, lines, and polygons will have a single unique user-id number. Figure 1 shows graphic examples of acceptable and unacceptable topological conditions.

**FIGURE 1:** Graphic examples of unacceptable and acceptable topological data conditions for digital data

<table>
<thead>
<tr>
<th>Topologically</th>
<th>Topologically</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unacceptable</strong></td>
<td><strong>Acceptable</strong></td>
</tr>
</tbody>
</table>

Overshoots

Slivers
3.7 Edge Matching

- Contiguous features should be edge matched or closed.

Line segments (arcs) that intersect the boundaries of a coverage must be accurately edge matched with the corresponding arcs in the adjacent coverages. Computer edge matching techniques ensure an exact match. In lieu of an exact match, arcs must be matched to within 0.01 inch, centerline to centerline.

- Closure lines should be attributed as “closure”.

In the absence of more accurate data, where edgematching line segments from adjacent coverages would move a feature or features a distance that exceeds NMAS standards for scale, the addition of a closure line is the preferred. A closure line is an arbitrary feature drawn to close a polygon or arc. Features that represent closure lines should be attributed accordingly.

- Map features should not to extend beyond prescribed dataset boundaries.

Arcs must not extend beyond (overshoot) nor fall short of (undershoot) the dataset boundary.
3.8 Coding Accuracy

- 99.5 percent of all attributes will be coded correctly.

The most serious problem is miscoding a feature. Coding an area of glacial till as a sand and gravel deposit is an example of this type of error. Other errors in attribute coding include any occurrence of misspelling or omissions. All attributes will be coded correctly.

3.9 Accuracy Assessment

The following assessments will measure data accuracy:

- 90 percent of all features within 0.01 inches when reproduced at the scale of original manuscript map
- 100 percent of all features within 0.02 inches when reproduced at the scale of original manuscript map
- Topology complete and accurate
- 100 percent of all attributes coded correctly

It is the responsibility of the producing agency to verify that the original data have been encoded within the accuracy limits set by these standards. Ninety percent of the cartographic features on a map will be digitized within 0.01 inches measured from the centerline or center point of a feature. One hundred percent of all cartographic features will be digitized within 0.02 inches.

There are many ways to formulate a scheme to assess data accuracy. Methodology to assess data accuracy should be developed with the automation contractor or within the producing agency if data are automated in house.

If all criteria are met, then note in the documentation (next section) that the digital map meets these standards.

3.10 Proof Plots

- Mylar proof plots
- Pen width - 0.01 inch

For each digital map created, a mylar proof plot of that map should be made to verify that the original data have been digitized within the accuracy requirements of these standards. The proof plot should be plotted on mylar at the same scale as the manuscript. All lines on the proof should use a linewidth of 0.01 inch or less and be drawn with liquid ink pens. Point data should be represented with the "+" symbol.

3.11 Attribute Coding

- Choose attribute coding schemes that are well defined and in common use.

- Names of new datasets should not be the names of existing datasets in the Maine GIS database.

- Begin all dataset names, item names, and attribute codes with a letter.

- Eliminate punctuation, i.e. hyphens, pound signs, periods from dataset names, item names and attribute codes.

- Limit the length of dataset names and item names to 8 characters or digits.
Define items according to intended use, i.e. numeric data types for statistical data or for data that will be used for calculation, characters data types where these functions are not anticipated.

Limit the length of attribute codes to 12 characters or digits.

Eliminate the use of the Boolean values 0 and 1 as attribute codes.

Wherever practicable utilize character attribute codes, i.e. previously defined acronyms or abbreviations in common use.

Naming conventions, item length, and punctuation recommendations are based on format conversion and compatibility requirements. For example, conversion to shapefile format changes some characters ( # - .) to underscores; item names over 8 places are truncated; the Boolean values 0 and 1 impact the results of geographic analysis. An effort will be made over time to consider these limitations in item names and to eliminate the use of these characters in data included in the Maine GIS database.

Use Standard Geocodes for Maine.

As an information service to state agencies and the public, Standard Geographic Codes for Maine, are available for download through the Maine Office of Geographic Information Systems (MEGIS) Data Catalog. The published table contains the first official Standard Geographic Code endorsed and adopted by the Governor of Maine, on July 1, 1971. Geocodes have undergone subsequent revisions, all of which were "officially" rolled back to this 1971 list by Maine's Information Services Policy Board (ISPB), as of January 2000. In 1971, all Maine state agencies were requested to implement these five digit geocodes, in agency information systems, to build a base of data for Maine's Minor Civil Divisions (MCDs) and to promote data sharing. Data coded by MCD may subsequently be summarized by any other type of region, administrative district, natural area, etc. The first two digits of the geocode represent the federal code (FIPS) for Maine counties, the remaining three digits uniquely identify each of Maine's MCDs and Reservations.

Prior to the development of important new data sets, a detailed series of data layer specific content standards that expand on existing standards should be completed. Attribute coding and content for all new data sets will expand on adopted standards and reflect coordination with federal and regional content standards. For example, coding of a soils database will follow the conventions of the USDA Soil Conservation Service. Nationally recognized formats also exist for coding of data layers such as wetlands and land use. When devising a new scheme, it is important to allow for as many different foreseeable uses of the data as possible. For example, in a land use coding scheme, it is preferable to have individual codes for different types of agricultural use (e.g. cropland and pasture) than a single code specifying agricultural use. The individual codes may be aggregated, but the single code cannot be broken into its components.

3.12 Database Attributes

Add and properly define a database field to store a standard geographic reference.

Standard geographic reference exist at all levels of government and in the private sector, for example the U.S. Board of Geographic Names responsible for the Geographic Names Information System (GNIS), the National Bureau of Standards, Federal Information Processing Codes (FIPS). Many Maine state agencies maintain standard geographic reference information relating to their programs. Examples include the Maine State Planning Office, Standard Geographic Codes for Maine Minor Civil Divisions (GEOCODES), Maine Bureau of Parks and Lands, Coastal Island Registry (CIREG), Maine Inland Fisheries and Wildlife Maine Information Display and Analysis System (MIDAS). An effort is ongoing to promote the standardization of database attributes by working with agencies who serve as primary sources for geographic references. To take full advantage of Maine's growing capability to process, analyze and
display geographic information, databases developed within state government, whose records relate to a geographic feature, should include a minimum of one field that stores a related standardized geographic reference from a primary source, such as:

- Standard Geocode for Maine Minor Civil Divisions
- Federal Information Processing Standard or FIPS Code
- Latitude and longitude to the nearest second
- Universal Transverse Mercator Eastings and Northings to 6 and 7 digits, respectively.
- Coastal Island Registry Number
- MIDAS Number
- Tax Map Identifier and Lot Identifier
- Street Address

Georeferenced databases can be linked in the GIS environment for analysis and display with the spatial data that makes up the Maine GIS database. Georeferences can be very specific, for example: latitude and longitude locations for storage tanks and hospitals. Some data, for example disease occurrences or demographic data, can be assigned a less specific georeference like a geocode or a census geographic unit. Many phenomena can be mapped including: bus routes and school bus stops, chemical and oil spills, contamination, hazardous material locations, storage tanks, administrative districts, emergency service locations, disease occurrences, demographic characteristics, clients, services, natural and biological resources.

4. FGDC DOCUMENTATION

- Federal Geographic Data Committee (FGDC) compliant documentation is required for all data to be included in the Maine GIS database.

The FGDC Content Standard for Digital Geospatial Metadata (CSDGM) has been adopted by the GIS EC as the documentation or "metadata" format for all data included in the Maine GIS database. Participating agencies are responsible for providing documentation in FGDC compliant metadata format for geospatial data products. The metadata must be stored and maintained by the agency developing and maintaining the geospatial data.

FGDC metadata is the required documentation of all geospatial data products produced through partnerships, grants or contracts with federal agencies. Federal agencies collecting or producing geospatial data, either directly or indirectly (e.g. through grants, partnerships, or contracts with other entities), are required to ensure, prior to obligating funds for such activities, that data will be collected in a manner that meets all relevant standards adopted through the FGDC process. (Executive Order 12906, published in the April 13, 1994, edition of the Federal Register, Volume 59, Number 71, pp. 17671-17674. COORDINATING GEOGRAPHIC DATA ACQUISITION AND ACCESS: THE NATIONAL SPATIAL DATA INFRASTRUCTURE)

Metadata is "data about data". Like a style guide, the FGDC CSDGM defines what information belongs in a metadata record and the order in which it is presented. The goals and objectives of Maine's FGDC compliant documentation can be summarized as follows:
1. to provide a history of each geospatial data set included in Maine’s GIS database with standardized information on content, location, purpose, accuracy, condition, quality, collection and development processes, scale, projection, feature attribution, and other characteristics of geospatial data necessary to determine the utility of a map for a specific purpose;

2. to protect Maine’s investment in geospatial data by minimizing the risk of data loss, and minimizing the recreation of existing data, through systematized documentation of geospatial information; and

3. to promote data sharing by providing information about Maine geospatial data holdings to external catalogues, clearinghouses, and brokerages and by providing information on the processing and interpretation of spatial data received through a transfer from an external source.

4.1 Basics of FGDC

FGDC compliant documentation or metadata tells you the who, what, why, how, when and where of geospatial data. The standard organizes a metadata record into seven main sections.

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

4.2 Technical Support for FGDC

Technical support for content development of FGDC metadata and information on FGDC compliant software for metadata development is available through the MEGIS, Data Center, Technical Support phone line at (207) 287-6144. A link to additional FGDC resources is available under “Standards and Guidelines”, MEGIS homepage, http://apollo.ogis.state.me.us.

4.3 More About FGDC

Maine FGDC compliant metadata is published through the MEGIS internet Data Catalog. FGDC metadata documents are also the basic components of the National Geospatial Data Clearinghouse, and FGDC compliant metadata for Maine’s geospatial data can be made available to clearinghouse nodes nationwide. The National Geospatial Data Clearinghouse is a distributed online catalog of digital spatial data, part of the National Spatial Data Infrastructure (NSDI) (Executive Order 12906). NSDI goals are to reduce duplication of effort among agencies, improve quality and reduce costs related to geographic information, to make geographic data more accessible to the public, to increase the benefits of using available data, and to establish key partnerships with states, counties, cities, tribal nations, academia and the private sector to increase data availability.
REFERENCES


University of Texas at Austin School of Architecture, 1999. National Map Accuracy Standards (NMAS) Horizontal Accuracy Examples. Community and Regional Planning Program, University of Texas at Austin, Austin, Texas.