cartographic techniques

Raster data in multimedia atlases: benefits and challenges. Examples from the "Atlas of Switzerland – interactive"

Marianne Ruegsegger,
Christoph Schmid and René Sieber
Institute of Cartography
ETH Zurich, Switzerland
e-mail: ruegsegg@karto.baug.ethz.ch, schmid@karto.baug.ethz.ch, sieber@karto.baug.ethz.ch

Introduction

Interactive atlases historically have not used raster data; instead vector data and graphics are used. It is important to note the difference between scanned maps, which fall into the category of graphics, and raster data. Raster data has attribute information that can be queried on a cell-by-cell basis—this is the most important quality raster data affords to an interactive atlas. Typically, this kind of querying of data has been the province of GIS software, but the richness of information and the ability to provide specific answers to questions about exact locations make raster data a logical choice for interactive atlases. Vector data values are typically pre-classified or have been averaged across areas that are much larger than the raster cells that were used to record the data.

The purpose of this article is to provide strategies for including raster data in digital atlases in order to create the richest experience possible for users. Possibilities and difficulties will be discussed. As the second edition of the "Atlas of Switzerland – interactive" will focus on environmental issues, environmental data in will underpin these discussions. Additionally, new tools that were created for, and examples from, the upcoming edition of the "Atlas of Switzerland" will be shown. Also attention will be paid to techniques for combining raster and vector data, as well as raster and raster data.

History of the Atlas of Switzerland

Two printed editions of the "Atlas of Switzerland" were published between 1965 and 1997. These contain nearly 600 individual maps. In 1995 the Swiss government planned a completely new edition, both a printed and a digital, to be published in four languages (German, French, Italian and English). The themes were mandated [Hurni et al. 1999]:

- Landscape and environment
- Population and society
- Economy
- Infrastructure and traffic
- Culture and politics
- International relations

The first interactive edition was published in the year 2000 as "Atlas of Switzerland – interactive" on CD-ROM and included 250 thematic maps on population and society. Additionally, statistical maps for Europe and a 3D topography tool that showed 3D relief maps, block diagrams and panoramic views were included.

The Atlas of Switzerland is currently being updated with new data and new functions; the most important of these are:

- Updated statistical data on society and population
- 400 new themes covering landscape and environment
- The design of the user interface and navigation tools
- Methods for query and display of data
- Substantial expansion of the use of 3D data

With new data the requirements changed; environmental data is rarely available as vector data, therefore raster data needed not only to be considered, but the challenge was really to find ways to include it.

Data in Multimedia Atlases

The Atlas of Switzerland team did not find any multimedia atlases that contained maps made with raster data, just vector data and graphics. Experience in working with geographical data sets showed raster data, especially in environmental research, is used at least as often as vector data. This leads to: why aren't raster data sets used in multimedia-atlases?

One idea is that traditional cartographic representation is a highly generalized and abstracted model of the real world. Traditional cartography is expected? Traditional cartographers did not have the benefit of an abundance of remotely sensed or photogrametric data. Interactive atlases offer an opportunity to break from the traditional and incorporate valuable data that logically belong as content.

Another response is that there are strong tendencies now, especially on the Internet, to replace raster graphics with "smaller, more accurate" vector data. This is an unfortunate confusion of among other things accuracy and coordinate precision and that vector data is smaller in terms of bytes to be transmitted. The "Atlas of Switzerland-Interactive" must show the best data possible and whether the data is raster or vector should not be an issue. In fact, if a map must include raster, or raster and vector data to be optimal, then that is what must be.

Still, there is some allure to the idea that it may make sense to vectorize raster data. For instance, the attributes of rasters of different resolutions can be generalized to polygons with nice presentable consistent edges. Beyond leading readers to falsely lofty assumptions about the data's accuracy, there are many disadvantages to this procedure. Most importantly it's a time consuming process that cannot be fully automated and therefore is very costly.

Comparing Raster to Vector Data

In this short review raster and vector data are compared (Table 1) and, advantages and disadvantages are listed (Table 2) and discussed.

The most important characteristic of vector data certainly is its topology; the data is "intelligent", redundancies are avoided and neighbors are known. Vector data consists of objects with clear borders that define regions where the data values assumed to be uniform. For many continuous phenomenon vector data representations are more abstracted than raster data. The information in raster data on the other hand is continuous and covers all space, which makes it suited for data modeling [Hohl 1998].

Data can be captured by digitizing (vectors) or scanning (raster). The latter is fast and therefore less expensive, digitizing, however, is a time consuming and highly repetitive job that requires skill and care. Once the data is digitized though, it's easy and inexpensive to update.

Raster data is less abstract than vector data. Depending on the cell size it is even possible to create realistic pictures of the landscape. Cells values can be, but do not need to be classified.

The big drawback of raster data is the storage space it needs, as every cell has information attached to it. Bigger cell sizes can reduce the storage need but result in less attractive maps. However, relative to vector data with many attributes, or that is highly detailed, raster data is no less difficult to store or transmit.

Looking at this synthesis, it's obvious, that it's not a raster vs. vector, but a side by side of raster and vector that will bring the maximum benefit. Raster data is excellent when using modeled datasets, as found often in environmental topics, where as vector data is optimal for the mapping of entities with clearly defined boarders such as lakes, streets, etc.

Raster	Vector
Neighborhood	Topology
Suited for data modeling (interpolation and comparative analysis)	Representational Accuracy
Stored as location (discrete units)	Stored as objects
Continuous geographic variation	Information within object seems to be uniform

Table 1. Characteristics of raster and vector data

Raster	Vector
Fast data capture (inexpensive)	Topology
Greater realism (less abstract)	Update easy
Overlays computationally faster [DeMers 2002]	Accurate
Variation greater (changes on a cell by cell basis	Easy to add attributes

Table 2. Advantages of raster and vector data

Raster	Vector
Storage	Data capture is time consuming
Not intelligent (no topology)	Expensive
Relative lack of spatial resolution	Storage
Bad for line representation	

Table 3. Disadvantages of raster and vector data

Raster	Vector
Good for environmental data, resource management, remote sensing	Good for mapping, surveying, utility information system
Fast but less precise	More graphical
Suited for complex modeling	Difficult to use for modeling
Interpolation of point data to surface	Symbolic, more cartographic

Table 4. Comparison between raster and vector data

Use of Raster Data in the "Atlas of Switzerland – interactive"

Given the above analysis of raster and vector data that the two would be used side by side in the upcoming edition of the "Atlas of Switzerland – interactive". Therefore the intent for the user interface was that it should remain unchanged and preferably all features and functionality for vector data should also be available for raster data--most importantly query functionality. There are a few features that are exclusively available for raster or vector data.

Features that allow little or no user interaction:

- Color interpolation: raster data sets are colored according to the cell values. Each value can be assigned an individual color, or a color can be assigned to a value-range. The colors are defined for the class bounds. If a cell value doesn't lie exactly on one of the class bounds, it can be colored with a mix of the two colors of its class bounds instead of being assigned to either one of the colors.

- Space interpolation: the viewing resolution of the displayed map is often different than the source raster data. This can be avoided by interpolating the map points (but not the underlying raster values), resulting in a consistent display.

- A "mute" layer can be added to active layer to embellish it: this layer usually consists of lines or points. This layer merely contains graphical objects that cannot be queried. It usually consists of line or point objects that enhance the meaning of the map. Fig 1 (page 83) shows the magnetic declination. The underlying raster layer can be queried, while the isogonic contours, the "mute" layer, cannot. But it helps the user to quickly make out areas of similar declination.

- 3D view of maps (See Fig 2, page 83): the DTM (digital terrain model) can be overlaid by a number of raster maps. Both, block diagram and panoramic views are possible. The maps can also be queried.

Following are the main functionalities for raster data that will be available to the user:

- Interactive query of cells
- Interactive change of color values and classification of cell values
- Overlay the current map with one of several pre-selected layers.

Method: Requirements for raster data

A multimedia atlas should be able to display data and produce interactive, aesthetically pleasing, screen maps. Therefore cartographic aspects are more important than analysis or GIS functions, e.g. the altering of the data. Thus we propose the following separation of analysis and visualization parameters within the data:

- Raw data: the discrete original values are stored in an appropriate format that meets the requirements given later on.
- Georeference: puts the data into its geographical context. This usually includes corner coordinates, cell size specifications, and coordinate system definition.
- Visualization definitions: needed to display the data on the screen. This can be done by a value classification and its class color definitions.

With this information we can produce a raster data map by adhering to the following formula:

Data + Visualization parameters = Map

It is essential to base maps on the original data and not any classified values or even color values (which in fact would be a graphic image). This way it is possible to incorporate any or many interpolation methods and create, for instance, a set of classified temperature maps where each still returns the true temperature values when a location is queried.

For storage efficiency and computational ease, data should be numeric. All non-numeric data can easily be converted to values that are later assigned the original text. This is because a potentially, a huge disadvantage of raster data is its size. For example, a raster with hectare-sized cells covering Switzerland with 4-byte integer values produces a raster of approximately 39Mb. This demonstrates that it is of utmost importance to reduce the data size both on disk and in the memory.

A simple and common way to reduce a file's size on disk is to compress it with one of the numerous compression algorithms available. Compressing the data is a good start, but it doesn't reduce the data size when loaded into memory.

In many cases the data's format space is over-specified, e.g. for a raster containing data values from 0 to 255 a single byte is sufficient to store a single value, but often 4 bytes are used. The appropriate data type for a raster dataset should be determined by analyzing the data it contains.

Another way to reduce the space the raster data uses in memory is to fragment the dataset into smaller tiles. This way only the tiles currently visible are loaded into the memory. The disadvantage of this technique is that the

Type	Value Range	Size per value
Floating point		4 bytes
Signed integer	-2'147'483'648...+2'147'483'648	4 bytes
Unsigned integer	0...+4'294'967'276	4 bytes
Signed short integer	-32'768...+32'768	2 bytes
Unsigned short integer	0...65'536	2 bytes
Byte	0...255	1 byte

Table 5. Data types and storage sizes

resulting grid has to be composed out of multiple tiles, which causes longer map composing times.

Clearly it does not make sense to directly use a raster with a 10m cell-size for a map with a 1km pixel resolution. Reducing the resolution 'on the fly' while generating the map is very time-consuming and usually ignores the data's properties as pointed out later. But if several rasters are calculated and stored in various resolutions in advance, loading time can be drastically reduced. When cutting the resolution in half step by step and storing all resolutions together 33% additional storage space is needed at most:

Reducing resolution has to be

$$\sum_{n=1}^{\infty} \frac{1}{2^{2n}} = \frac{4}{3}$$

done with care as the data's meaning could be compromised. It is no problem to resample a grid containing continuous data by using common interpolation techniques (nearest neighbor, bilinear interpolation etc.). With nominal data, resampling methods have to be used that preserve the data values (e.g. it's not possible to interpolate between water and agricultural use in a land use data grid).

These considerations show, that, to assure correct data resampling and to increase performance, it is mandatory to preprocess the data and to provide it in different resolutions.

Results: The Implementation

For the this edition of the "Atlas of Switzerland" a proprietary grid format, AoS Grid Format was created that meets the specific requirements of the atlas. It contains consecutive raster data of a numerical data type. It also contains all necessary header information like the georeference, cell sizes, data type etc.. The data block of the file can be compressed to reduce the file size and encrypted to protect the data. This file format is designed to contain a single grid in a single resolution.

For performance reasons the grid was pre-calculated in different resolutions and cut into tiles. Hence a "AoS Tile File" format was created to facilitate merging these multiple grid tiles and multiple resolutions. Through the header information this file provides direct access to any grid tile in its various resolutions.

The simplest use of the AoS Tile File is to store a single untiled grid in one resolution. In this case just some metadata is added (e.g. the file description string). In more complex cases the file contains the grid in multiple resolutions with every resolution tiled into any number of tiles. The different resolution layers even can have their own georeference.

The tile file format is designed to contain various kinds of data, not just grids. For example it can also contain an image or a digital elevation model.

Visualization

The Atlas of Switzerland software engine makes the final map. The required grid tiles are loaded and merged together, then the visualization definitions are used to assign every grid value a color. The resulting image must be resampled to fit the final display area. This resampling is delicate since it could change the map appearance unintentionally (e.g. by interpolating between the pixels).

The following example (Fig. 4, page 84) shows precipitation. The raster data has partially transparent vector data (gauging stations) draped on top. This shows how two sets of data and their respective visualization parameters have been combined to create one map.

Conclusion

For the new edition of the "Atlas of Switzerland – interactive" environmental data, typically stored in a

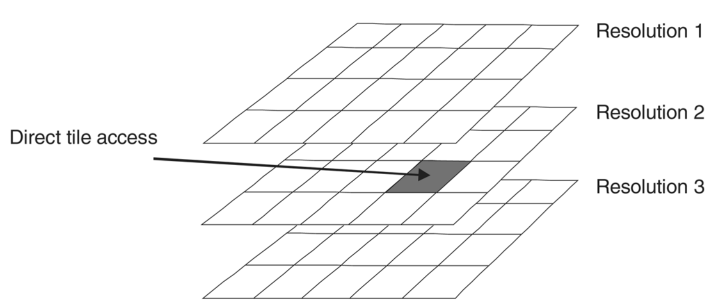

Figure 3. Data access in the AoS Tile Format

raster format, will be used. Raster data doesn't have to be vectorized, and it makes a great contribution to a multimedia atlas.

The "Atlas of Switzerland – interactive" uses raster data and fulfills the main requirement of an interactive atlas: that the data can be queried. In the "Atlas of Switzerland – interactive" continuous information can be queried on a cell by cell basis, returning unclassified values. Raster data is used side-by-side with vector data sharing the same user interface.

Fast data access is possible, mainly enabled by the use of an appropriate data model. The main characteristics of this data model are:

- Raw, numeric data is used with additional specifications given in description files

- Each data set is preprocessed and stored in proprietary file formats:
    - The appropriate data type (float, signed integer, byte, etc) is individually assigned. The data is stored in the AoS grid format.
    - Big raster data sets are resampled into several resolutions and fragmented into various tiles. Based on a map's scale, only the matching resolutions and tiles are loaded and thus memory use and loading time are significantly reduced.

- The final visualization is created by using information stored in separate description files.

References

DeMers, M.N. 2002. *GIS modeling in Raster*. John Wiley & Sons.

Hohl, P. Ed. 1998. *GIS Data Conversion: Strategies, Techniques, and Management*. Santa Fe: OnWord Press.

Hurni, L., Bär, H.-R., Sieber, R. 1999. "The Atlas of Switzerland as an Interactive Multimedia Atlas". *Multimedia Cartography*: 99-112. Berlin Heidelberg: Springer-Verlag.

Tomlin, C. Dana. 1990. *Geographic Information Systems and Cartographic Modeling*. Englewood Cliffs: Prentice-Hall.