

# PRODUCING MILLIONS OF MAPS FOR THE UNITED STATES' 2010 DECENTNIAL CENSUS\*

M. Schell<sup>a</sup>, S. Spahlinger<sup>b</sup>

<sup>a</sup> U.S. Census Bureau, 4700 Silver Hill Rd, Washington DC 20233-7400 – matthew.c.schell@census.gov

<sup>b</sup> U.S. Census Bureau, 4700 Silver Hill Rd, Washington DC 20233-7400 – stephanie.m.spahlinger@census.gov

\* This paper reports the results of work undertaken by U.S. Census Bureau staff. This paper is released to inform interested parties of research and to encourage discussion. Any views expressed are those of the author and not necessarily those of the U.S. Census Bureau.

**KEY WORDS:** Census, automated mapping

## ABSTRACT:

The U.S. Census Bureau designed and produced millions of unique paper maps to support approximately thirty different census operations for the 2010 Census. Many operations were short term and a few involved over 100,000 temporary enumerators. A compressed operational schedule, coupled with high volume map production, required a non-interactive mapping process. In this environment, a computer performs all cartographic functions. In its customized mapping system, the Census Bureau employs automated computer algorithms that perform traditional cartographic functions such as scale determination, feature density assessment, inset area identification, map sheet layout configuration, text placement, integrated quality control, parameter driven symbolization, and specifically formatted output. This paper describes the Census Bureau's mapping goals, the rationale for the current approach, and a description of the map production system and methodology.

## 1. INTRODUCTION

The United States Census Bureau's Geography Division produced more than 10 million small format maps and more than 4 million large format maps to support the 2010 decennial census field operations. The Census produced these maps, almost all for hard copy output on physical paper, using a digital, 100% automated map creation system, the Census Automated Map Production System, or CAMPS.

In this written paper we will detail the internal workflow of that automated system. While it's highly unlikely that in the future any organization will claim in these proceedings to have produced almost 15 million unique maps for old-fashioned paper output, the automated cartographic logic built in to CAMPS is at its core relevant to all kinds of automated map producing systems, be they on the internet or in the palm of your hand.

### 1.1 Rationale and Goals

Following the 2000 census the US Census Bureau migrated from a custom-built, in-house spatial database called TIGER, to a commercial Oracle Spatial database now called MAF/TIGER because it also incorporates the Census Bureau's Master Address File (MAF). Census cartographers designed a new in-house system, CAMPS, to create maps using the MAF/TIGER database in Oracle.

In the lead up to the census-taking itself, the majority of the maps produced by the Census were used internally, to assist census workers in delegating workloads, locating

assignment areas, and recording census data. With the completion of census taking in 2010, we are using the same system to create maps to publish data for public use.

Creating this many maps for a variety of users and uses requires two types of resources, and as with all resources, neither are in infinite supply. First, producing close to 15 million maps using ink and paper requires not surprisingly both ink and paper, as well as other tangible hardware resources. Second, the people and time available to produce these maps are also finite, and we could hardly expect a staff numbering in the tens to produce a volume of map sheets numbering in the millions with even a minimal level of human interaction with the maps. Consequently, one of the Census' overarching goals in map production is to produce the most usable maps possible on the fewest map sheets.

These two constraints suggest the need for a system that is maximally automated, while at the same time able to produce high quality maps. In the absence of available commercial software that met all of our requirements, the Census Bureau produced its own system. The Geography Division at the Census Bureau has a long history of producing such systems (Trainor 1990); (Beard and Robbins 1990).

## 2. METHODOLOGY

### 2.1 Parameter Driven

CAMPS attempts to mimic the process that a cartographer would follow when creating a map by encapsulating all possible decisions in pre-populated parameter tables. Parameters are logically grouped into several database tables related to the map content they describe, with records for a single set of mapping parameters (called a project) linked from table to table by unique identifiers, or keys. The following is a list of the parameter table names and a brief summary of each table.

**Parameters** – Basic map information, for ex., project code.

**Canvas** – Index, parent, and inset, each with separate rules.

**Sheeting** – Description of sheets, for example height and width.

**Layers** – What will appear on each canvas/sheet combination.

**Symbology** – Named symbology for each layer.

**Styles** – Style for each symbology, for example color and font.

**Definitions** – Definition for each style, for ex., color values.

**Marginalia** – Description of marginal elements.

**Postprocess** – Description of supplemental files and embedded properties.

In order for the system to be entirely automated, the parameters describe everything from basic information about where in the database to source the map subject to very low-level details like specific color definitions and fonts.

PROJECT_CODE	CANVAS_LIST	GEOMETRY_TYPE	GEOMETRY_CLASS	FEATURE_DEFINITION	STYLE_DEFINITION
1245	INDEX	AREF	COLLBLOCK	F A	*pgX
1245	INDEX	AREF	GLACIER	S A	*pgX
1245	INDEX	AREF	HYDROA	I A	*pgX

Figure 1. Project Symbology Table

### 2.2 Scaling

When the processing to create a map for a single entity begins (for example a county), the first step is to determine a map scale. This scale then drives all of the following layout decisions.

Census cartographers choose from several algorithms available to determine a scale, basing the decision on the primary purpose of the map and what elements must be labelled for the map to be effective. For example, if the purpose of the map is to clearly label census blocks then the system will use an algorithm that takes the census blocks within the map extent and calculates a scale where some percent of virtual, not yet mapped, block labels will fit within the area extent of those blocks.

When a mapped subject contains many small, densely packed blocks, like in an urban area, the scale will be larger than a map of more rural areas. Similar algorithms can be used to drive scaling based on labelling linear features or points.

Imagine continuing with the example of scaling to census blocks. For any block that will be labelled, the cartographer knows the size in square inches,  $S$ , of a typical block label at a desired font. Calculating  $S$  doesn't require anything fancy. Printing a bunch of "8's" at the desired font and fetching a ruler will usually work well enough.

The cartographer also knows that if every block was entirely covered by its label, things would be really crowded. Only some percent of the block,  $P$ , should be covered by  $S$ . For any one census block, which has a size on the earth recorded in the database, converted to square inches of area  $A$ , the desired scale for that block is

$$Scale = \sqrt{\frac{A * (P / 100)}{S}} \quad (1)$$

The mapping system then takes the desirable scale for each block on the map, sorts the scales, and picks the scale at some percentile of the list. Blocks that fall outside of this percentile may not be able to be labelled easily inside the block itself without resorting to leader lines or a smaller font. If there is a cluster of these dense blocks they may also appear on a separate inset sheet.

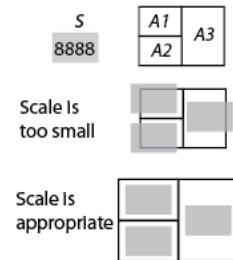


Figure 2. Scaling Schematic

If the scale determined is small enough that the entity being mapped can fit on a single map sheet, the system overrides the calculated scale and determines a scale that will fit the entity snugly on the sheet. For a mapped entity with a minimum bounding rectangle of height  $H$  and width  $W$  in ground units, and a map sheet of height  $h$  and width  $w$  where  $h$  and  $w$  have been converted from page inches to ground units, the best scale to fit an entity to a sheet is based on the constraining dimension, height or width.

(2) 
$$\begin{aligned} & \text{If } (H / W) > (h / w) \\ & \quad \text{Scale} = H / h \\ & \text{Else} \\ & \quad \text{Scale} = W / w \end{aligned}$$

### 2.3 Inset Determination

With a scale chosen, the software next determines whether to produce any insets. The software finds clusters of dense

features based on polygon or point feature density. Again, the parameters that drive inset determination are set by the cartographer in advance.

The insetting step finds dense clusters using the DBSCAN algorithm (Ester et al., 1996) by centering an imaginary box, converted from page units to ground units at the given scale, over the same layer used in the initial scale calculation, and counts the number of features that fall within the box. If there are more than a specified number of features then that area is a candidate for an inset.

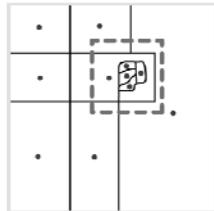


Figure 3. Inset schematic. The centroids of the census blocks inside the dashed box may be a candidate for an inset area.

#### 2.4 Sheeting

With a scale and insets determined, the next step is to decide, for maps that require more than one sheet, an optimal sheet configuration. CAMPS tests 16 different configurations, draping virtual sheets at the chosen scale over the subject entity. The software chooses the configuration that requires the fewest sheets.

Figure 4 depicts two of the sixteen configurations, “upper left,” and “centered out.” The upper left configuration in this example requires 6 sheets compared to the centered out configuration’s 9.

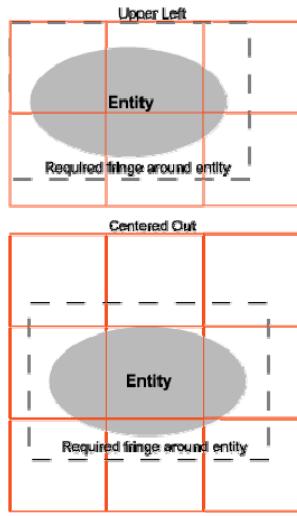


Figure 4. Sheet configuration

#### 2.5 Iteration

At this point the system has determined a single acceptable scale and sheet configuration. However, as mentioned in

the introduction, one of the driving motivations in the system is minimizing the number of sheets. To this end, the system next iterates through the same process all over again using a larger scale, with the list of available scales determined in advance by the cartographer.

A scale larger than previous scale already tested will allow labels to fit better. And if the resulting sheet configuration produces fewer sheets, reducing insets, then the system chooses the larger scale as the preferable one.

#### 2.6 Geometry Gathering and Normalization

With the sheet configuration locked down, the system now knows the maximum extent, on the ground, of the area that it will be mapping. The next step is to gather and normalize into local database tables all necessary data to produce the current map. The advantages of moving all data into local tables are two-fold.

First, the MAF/TIGER database is spatially large in extent and contains many vintages (or versions) of data for any given spatial extent. In subsequent geometry processing the mapping system will be repeatedly gathering layers into memory for processing, and to be constantly sifting through a large national dataset to select out just the data of interest impedes performance. Selecting from a local temporary data cache is more efficient.

Second, data in the MAF/TIGER database is generally not normalized for mapping purposes, but rather for warehouse functions such as update and maintenance. As a result, even simple database queries like “what is the name of this feature?” are not easily extracted from one layer or table to the next. Before continuing with further cartographic processing, it’s useful to bring all of the data to be mapped into local structures with standardized attributes like “name,” “object id,” etc.

#### 2.7 Geometry Processing

With the sheet configuration determined and all geometries gathered into local data structures, the system processes through the layers that the cartographer has selected to appear on the map. Each layer can be clipped and symbolized differently depending on where it appears on the map. The three symbologies that CAMPS creates are “subject,” “fringe,” and “inset.”

For example, on sheet number one of several sheets depicting a county, the area inside the subject county is “subject.” The area outside the county is “fringe” with respect to the subject county. And if there is a cluster of dense features requiring an inset, the space inside the inset area can also be symbolized differently, as “inset.” Any single feature can be clipped into one, two, or three of these symbologies, depending on which areas it crosses on the sheet.

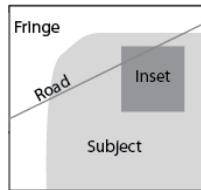


Figure 5. Subject, Fringe, Inset framework for one sheet

For example, in Figure 5 above a road may cross into and out of all three symbolologies. The system will create five separate output features from the one input road, where (in this case) two outputs are fringe, two are subject, and one is inset.

These multiple symbolologies are useful for cartographic design. For example, roads may be labelled in the subject but not in the fringe, or shown with a different line weight and color. And often features that cross into inset areas are not depicted at all, since the sheet where they are best shown and labelled is the inset sheet itself.

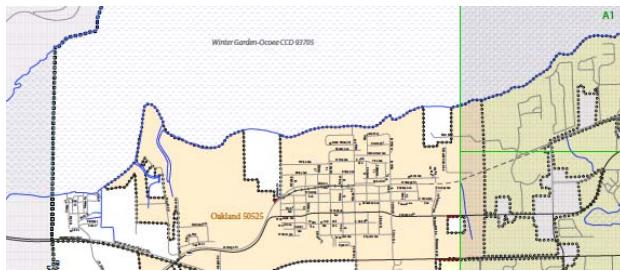


Figure 6. A portion of a census county-based map with fringe symbology on the left, subject in the center, and inset on the right.

A second major component of geometry processing, beyond clipping into symbolologies, is chaining and symbolizing boundary features. The Census Bureau maintains geographic entities in the MAF/TIGER database in a topological hierarchy, where higher levels in the topology, such as states, can be built directly from the same primitive nodes, edges, and faces as lower-level entities like counties. As a result, any given boundary line on the map may in fact represent the boundary of several entities in the census hierarchy.

In order to depict all of these unique geographies the census mapping system builds, for each primitive edge, a list of all geographies depicted on the map that share that edge. Then the edges are chained together, and the system draws the corresponding symbols, alternating between the symbols along the chains.

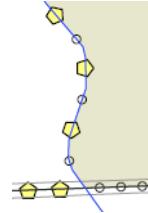


Figure 7. Alternating boundary symbology. A minor civil division is on the left, sharing a boundary with an incorporated place on the right.

In the final phase of geometry processing, the system cycles through several parameters that the cartographer has selected to improve the cartographic design. These represent “finishing touches” that a cartographer might make when designing a map by hand, but which can also be processed in an automated fashion. We highlight a few below.

**Label Largest Polygon:** For features with multiple pieces on the sheet, delete the name from all but the largest piece.

**Suppress Labels At Sheet Edge:** For features that cross onto another sheet, suppress labels at the sheet edge for small features.

**Add Characters to Features that Cross Sheets:** It’s often useful for a map user to know that a particular feature also appears on another sheet. The system adds a character, like an asterisk to indicate this.

**Delete Small Polygons:** Some regions are flush with tiny polygon features like lakes. If they are so small as to be almost invisible at scale, they can be deleted.

**Scrub Duplicate Fringe Labels:** Sometimes features will cross from subject to fringe multiple times and the fringe labels can be removed to reduce clutter.

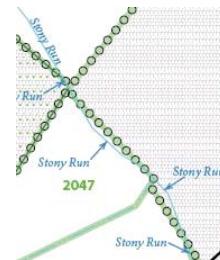


Figure 8. The Stony Run river does not need to be labelled 5 times.

## 2.8 From Database to ASCII

CAMPS next downloads the finalized geometry for each sheet to an in-house ASCII file type called the Map Image Metafile, or MIM. The MIM is a flat file of ASCII characters that is a self-documenting, full-image description of a map (Geography Division 1998).

Here’s a sample set of MIM tags describing a road with the name “Main St.”

```
*bef LINF_ROAD_S_447
*str 4 LINF_ROAD_S_447 Xadd
11.334 12.528 11.512 12.527 12.047 12.527 12.225 12.527
```

```
*vtx 11.334 12.528 00.097 0 LINF_ROAD_S_447 Xadd
"Main St"
*enf LINF_ROAD_S_447
```

The geometry for the road is described with four sets of coordinates in inches (11.334 ... 12.527), and the name of the road is fixed at a point near the beginning of the road (11.334 12.528). Higher up in the MIM, before the appearance of this set of tags, there would also appear several tags describing the font, line weight, color, and so on for the road and its label. This generic map description is human readable (in a literal sense), and can be easily converted to any other page description language.

The system calculates the conversion from real world coordinates in the database to map page units in the MIM using a standard conversion. For any X and Y coordinate pair in the database, that coordinate pair is offset from the lower left origin of all elements that will appear on the sheet, x and y, also in real-world coordinates. We also must calculate the conversion C from real world coordinates, like meters, to page units, like inches. Finally, we have an offset height and width, H and W, location on the printed page that is the origin of the lower left corner of the map window, since the map is unlikely to be placed at the extreme bottom left of the page.

$$\begin{aligned} \text{page\_X} &:= (((\text{X} - \text{x}) * \text{C}) / \text{scale}) + \text{W} \\ \text{page\_Y} &:= (((\text{Y} - \text{y}) * \text{C}) / \text{scale}) + \text{H} \end{aligned} \quad (3)$$

## 2.9 Text Placement

With the map image now fully described in the MIM, in the next phase the system determines label placement. Clearly labelling features is one of the most challenging aspects of automated cartography.

The Census Bureau, partnering with MapText Inc., has for years been a leader in the development of text placement software. This software uses the coordinates in the MIM to analyze the white space available on the map, and based on parameters entered by the cartographer in advance, the software cycles through placement options. The decisions made by the software, like moving the text to different locations, reducing the font size, or adding leader lines, closely mimic the decisions made when a human cartographer places labels interactively.

## 2.10 Marginalia

Placing the labels completes the inside-the-neatline portion of map creation. The software now has all of the information it needs to create marginal elements in the space between the neatline and the frameline, such as a scale bar, north arrow, and legend.

## 2.11 Postscript and Metadata

With all elements of the map, inside and outside the neatline finalized, CAMPS converts the ASCII MIM to

Adobe PostScript (also an ASCII format) and then from PostScript to PDF.

Throughout the process the mapping system captures metadata like scale, projection, and date produced to describe each map and map sheet. The Census Bureau uses this metadata to track production and control distribution and map printing. Selected metadata is also published with maps distributed to the public.

## 3. CONCLUSION

Producing many maps in a short time frame has become the hallmark of a successful census. In its customized mapping software the U.S. Census Bureau has created a fully automated system that produces high quality maps without any interactive editing.

Many of the steps we have described could be applied to building a map production system in any environment, not just in the context of an enterprise-level GIS database. As the Census Bureau begins to consider the 2020 Census, we likely will attempt to build on or re-use algorithms from the existing system, even if the end product changes with new trends in map use.

## REFERENCES

Beard, C., and Robbins, A. 1990. Scale determination and inset selection within a totally automated map production system. *Cartography and Geographic Information Systems*, 17(1), pp. 57-68.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining*, Portland, OR.

Geography Division, US Census Bureau, 1998. Map Image Metafile (MIM) Language Manual, Washington, DC. <http://www.census.gov/geo/www/mim/concept.html> (accessed 10 Sep. 2010)

Trainor, T. , 1990. Fully automated cartography: A major transition at the Census Bureau. *Cartography and Geographic Information Systems*, 17(1), pp. 27-38.