Complex Automated Map Production Workflow Supporting the 2010 Decennial Census

Timothy Trainor and Stephanie Spahlinger

ABSTRACT: The early AutoCarto meetings presented various new ideas for automating a conventional mapping process. Goals included an improved and efficient production process. Initial attention principally centered on automating production of map series with the underlying assumption that similar development efforts could be applied to other large mapping operations such as that experienced during the U.S. Census Bureau's decennial census. As this effort progressed, the notion of designing, implementing and using a map specific "cartographic" database evolved into a multi-purpose resource reflected in today's concept of a geospatial database.

A decennial census requires an efficient map production process to meet a demanding field operations schedule. The Census Bureau uses a variety of map types to collect data and administer those field operations. Huge numbers of maps are needed in order to portray the nation at the level of detail required for census work. Staging this enormous task is only possible with automated map production processes.

Map production work at the Census Bureau includes management of multiple databases, map design, intelligent automated map file creation, movement of files and paper, printing, and digitizing. Each of these components requires quality assurance and close management oversight.

This paper describes challenges in designing, developing, and deploying a complex automated map production workflow.

KEYWORDS: census map, map production, MAF/TIGER, 2010 census

Introduction

Conducting a national census is the largest of peacetime activities undertaken by government. In the U.S., the decennial census is mandated by the U.S. Constitution. Census results have a direct impact on apportionment of Representatives to the U.S. Congress. It requires a large temporary workforce to conduct and manage the activity. The workforce numbers in the hundreds of thousands. The temporary staff is assigned in nearby communities and oftentimes works for only a few weeks at a time.

A decennial census requires detailed maps of the entire nation to support a variety of field operations. Many distinct map types are needed for activities and programs leading up to the enumeration of the population. The 2010 decennial census includes multiple field operations as well as several geographic partnership programs with distinct participant constituencies such as local governments. Maps are used for most of these operations and programs.

Sources for Mapping

The early AutoCarto conferences attempted to find solutions for automating the conventional mapping process in response to the growing cost of labor. There was great interest in reducing the amount of time it took to produce a conventional map. The most efficient targets of this endeavor were map series where common elements could be programmed such as map dimensions, content, scale, standard symbolization, and referencing systems.

Preparing data for automated mapping was most efficient through some form of cartographic database, unlike the standard relational database structure at the time. Mapping characteristics and functions such as placement of text, symbolized content, and color were stored in the database, extracted, and then translated on an output device. Elements such as text font did not exist as a stored element or object. To create the likeness of a font, a series of vectors were drawn to emulate font characteristics, for example its style and size. While today these approaches seem crude, they were effective in uncovering areas for improvement.

The limitations of this approach were quickly realized. Capturing map data was an expensive endeavor. The cartographic database had one purpose; to store data for making maps. Storage of the data was preprogrammed as part of an established production process. For example, a name that identified a feature on a map was stored with its coordinate extent. Letters were converted to numeric codes representing each character for drawing the label without relationship to its feature. This single purpose enterprise exposed the need for a different direction.

A multi-purpose geographic database that supported early renditions of geographic information systems (GIS) was introduced. GIS components provided geographers, and eventually, most professions with data and tools to store useful information and to conduct analysis for decision-making, research, and other applications. Inevitably, a map was desired to visualize results of that work. Making a map in GIS was possible with a variety of highly interactive tools. The ability to produce many maps or different maps with complex rules was a difficult challenge.

Over time, different geographic data models have emerged to support a variety of uses. Data models designed for analysis may not be well suited to cartographic applications. Choices in a data model usually mean that geographic data require some form of data preparation for mapping. For example, if segments of a road are stored individually due to differences in attribution, applying a consistent linear patterned symbol may require chaining the segments together from beginning to end. This need can be addressed in the mapping application, however, if multiple uses of the same function are needed for other mapping, then it is likely more efficient if stored once in the source geographic database. As the notion of a geographic database evolved, real world geographic data was used in newer geospatial database constructs that supported multipurpose applications including mapping.

Mapping Process Determinants: Volume and Schedule

The use of a standard-scale map series for census operations is not effective. Maps that support a national census require a variety of map scales since the geography can differ from one place to another, even within a small area. Managers try to divide the work among enumerators based on the number of housing units they need to visit, so the size and shape of the assignments varies widely. Scale is driven by the purpose of the map. Maps that will be annotated by enumerators are very large-scale. Slightly smaller scales are used on maps that help orient the enumerator to their assignment area. Even smaller scales are required for maps used by a crew leader to manage the work of several enumerators.

In the 2010 Decennial Census, the lowest level of geography delineated to conduct the census is the census collection block. The greatest volume of maps produced for the national census is the census block map. This map portrays a single census block, a polygonal area of land usually bounded by roads or other visible physical features. Block maps are printed on tabloid size paper (11 x 17 inches) so they can be easily managed in the field. Enumerators use this map to find housing units. They also update the map with changes to the location or names of geographic features for subsequent digitizing into the MAF/TIGER database (MTDB)[1]. The scale of the map must allow enough white space on the map for the updates to be drawn clearly.

Two factors prevent any possibility of an interactive mapping process: the number of maps and the schedule of operations. Creating large-scale maps for the entire nation produces a huge number of map files. Printed maps require equipment and supplies and sufficient time to do the work. Field operations have specific beginning and ending dates. Costs of operations are closely controlled when employing hundreds of thousands of field workers for short durations. All of the steps leading up to each sequential field operation are closely monitored.

The country is divided into approximately 7 million collection blocks. The number of pages per collection block map varies depending on the complexity of the geography. Some blocks have a single map sheet. Other blocks with numerous internal features, a dense network of housing units, or an elongated shape require multiple sheets, sometimes including inset sheets with larger-scale depictions of particularly dense areas.

It is impossible to know the total number of map sheets required to support the national census in advance. Based on results from testing operations, cartographers estimate an average of three map sheets per collection block. The estimates are needed for different purposes such as determining computer processing requirements, calculating peak workloads, acquiring sufficient numbers of printers, and procuring supplies (paper and ink).

The volume of maps for the decennial census dictates that batch computer processes are essential for creating digital map files that make their way to decentralized locations for printing and use. Racks of CPU processors and terabytes of file storage are required during a three-year period of time beginning two years prior to census day. During months-long production periods, tens of millions of map files are created. With this type of production demand and the volumes of maps required, interactive mapping is not feasible.

Various field operations have specific schedules to assure that the census is taken and initial results are provided to the President by the last day of the census year.[2] While the size and scope of operations are enormous and their geographic extent is national, the importance of staying on schedule is paramount to a successful census. Once field staff are hired, they are trained and then deployed to their assignment areas where they use their maps. Not having maps produced and available on time leaves a paid enumerator with nothing to do.

The example schedule chart below (Figure 1) is a schematic of activities and steps that occur in preparing the MTDB processes with updates of address and spatial data. Planning charts like this are updated and monitored daily to assure that preceding,

---

[1] MAF/TIGER is the U.S. Census Bureau's geospatial database. It stands for Master Address File/Topologically Integrated Geographic Encoding and Referencing.

[2] See www.census.gov for more information on the decennial census

dependent activities are documented, that status of current activities are on schedule, and that products are delivered on time. These planning charts help to assure data quality and consistency, creation of derivative databases for product generation, and the creation of diverse address and spatial products including maps.

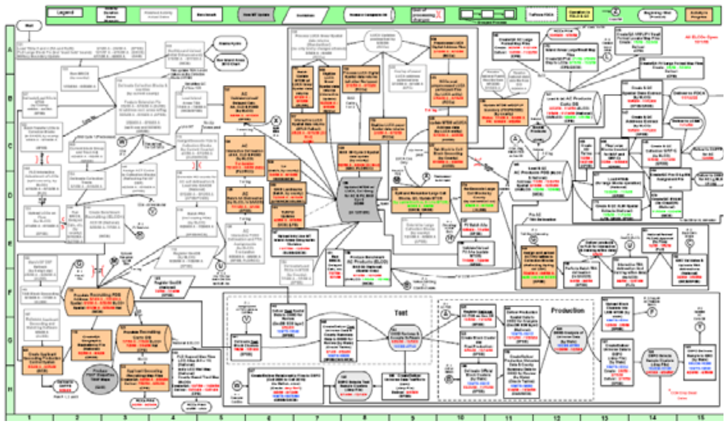

Figure 1. One sheet of a multi-sheet schematic document depicting MTDB update schedule.

Each geometric form in the chart indicates movement of data, a programming process, various forms of databases, and a variety of electronic products. Updates from partnership programs, field operations, and external data sources feed the MTDB. These are needed to support the next field operation and are reflected on the next version of maps. Voluntary nationwide partnership programs require careful coordination. Experience proves that participants and field staff oftentimes submit their work toward the end of a scheduled cycle, which does not match the planned even workflows and sometimes causes bottlenecks in the subsequent production steps. This is one of the reasons why monitoring is so critical. It allows for planning corrective actions.

Gathering Map Requirements

Cartographers begin by gathering requirements from sponsors of field operations that need maps. Completion of a map request form initiates the process (Figure 2).

Sponsor Name: ________________ Contact Information: __________

Purpose of maps: ______________ Context of Use: ______________

Type of Job Request
- ☐ Single map
- ☐ Small series (<100)
- ☐ Large series(>100)
- ☐ Unsure

Contain Title 13, restricted, confidential data?
- ☐ Yes
- ☐ No
- ☐ Unsure

Subject entity type?
- ☐ Nation (U.S. based)
- ☐ Census Region
- ☐ Census Division
- ☐ State
- etc.

Page Size
- ☐ Letter size
- ☐ Legal size
- ☐ Tabloid size
- ☐ Wall size

Desired Output
- ☐ Electronic (pdf, etc.)
- ☐ Paper copies
- ☐ Unsure
- ☐ Other

Delivery Date: ____________

Additional Information: ______________________________

Figure 2. Map request form

The requirements specify a variety of information such as map content, the size of the job, geographic coverage and the form of the output. Additional requirements may indicate that a map is used in more than one field operation or that updates are collected by enumerators and digitized into the MTDB. Census cartographers meet with the sponsors to discuss how the maps will be used and are told if any special requirements exist, such as data that is confidential which requires explicit notes on the map to assure special handling by staff. Arriving at a final map design is a process of negotiation that balances the sponsor's objectives with data and production constraints, and the cartographer's professional judgment. Often, several prototypes are created, reviewed, and revised in the process.

Map prototypes simulate the content and design for each map type. If an existing mapping system is in place then the prototype is created from it. If not, other software is employed such as ESRI's ArcGIS[3] to emulate the map design that ultimately will be developed within the Census Bureau's mapping system. The sponsor oftentimes shares the prototype with end users and adjustments are made based on responses.

The Cartographic Database

The MTDB is a seamless geospatial database. It covers the United States, Puerto Rico and the U.S. protectorate island areas. The database contains all of the geographic information needed to support the U.S. Census Bureau's censuses and surveys. Information in the MTDB includes features and content such as addresses, roads, rivers and water bodies, legal, statistical, and administrative boundaries, and their attributes such as codes and names.

Recently, the MTDB modernized from its original internally developed data structure to an Oracle platform. The specific version is Oracle 10G with Topology Manager. Assuring control of the source data in MTDB is paramount. Access to MTDB (referred to as the transaction database) is limited to staff performing updates, those maintaining software for the underlying core functions, and database administrators.

Daily updates are made to the MTDB. Products have requirements for specific geographic area vintage, for updates from geographic partnership programs or from field operation updates. A copy of the MTDB is made to assure that use of data reflects the requirements and prevents unintended consequences from a continually changing database. This snapshot version is called a "benchmark" of the MTDB.

The data in the MTDB is organized in its most elemental form. Geoprocessing efficiency and management are guiding principles. Use of the data is more effectively offered through a derivative database referred to as the product database where electronic extract products are created. Address extracts, geographic reference files and a variety of spatial data products such as shapefiles are examples of the diversity of offerings.

Mapping requires additional geoprocessing of MTDB data. A separate cartographic database creates additional Oracle tables from the benchmark that extend, reorganize, and normalize specific data to reduce duplication and excessive processing and is used to optimize mapping applications. For example, the cartographic database chains like-data based on attribute codes and/or names and the mapping application applies the specified map symbol to the chained feature. The existence of the cartographic database saves valuable time in meeting the difficult demands of the census schedule.

---

[3] The use of brand names does not represent an endorsement of a company or its products by the U.S. government..

Census Automated Map Production System

The Census Automated Map Production System (CAMPS) is the application that creates map files that meet requirements for a census operation. The cartographer uses the detailed map requirements to define a mapping project. Each map project is comprised of a number of separate components that, when combined, form a mapping system flow (Figure 3).

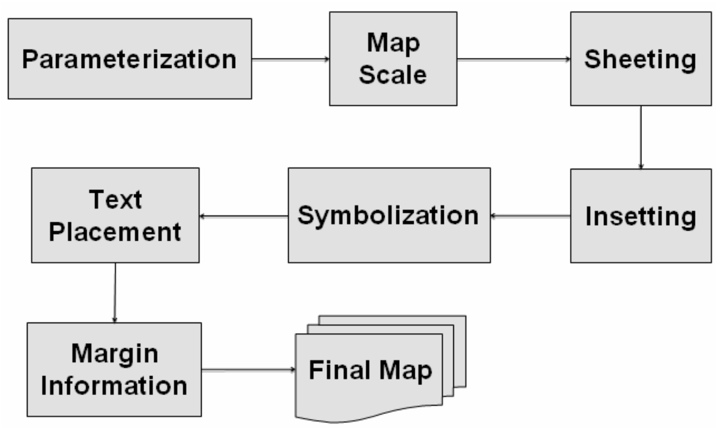

Figure 3. CAMPS mapping system flow

CAMPS components use different parameters to establish automated preferences throughout the mapping process. Some parameters reflect minimum requirements of a map project. Others offer a range of choices that result in better designed maps. Parameters are stored in Oracle tables that identify the content of the map, symbolization of the content, map image area and sheet dimensions, rules for scaling and insetting, rules for text placement, order for drawing content, and margin content and rules. Parameters influence each of the CAMPS modules below.

Map usability drives scale determination. Scaling parameters vary depending on user requirements. In some cases, the user only needs to see the information on the map while in other cases, the user requires adequate white space to apply feature and attribute updates to the map that will be digitized into the MTDB. In the former case, the ability to automatically place names on the map may drive the scale whereas, in the latter case, the density of features may determine the scale. In either case, a methodology is selected and the conditions are preset as part of the parameters in Oracle tables.

Sheeting occurs next in the mapping system flow. Some map types only allow a single sheet. In most instances, there is no restriction on use of multiple sheets which tend to be the norm. A guiding principle in sheeting is to produce the minimum number of sheets. This principle simplifies field use and minimizes the user's interaction with multiple sheets.

Sheeting and scaling for census use are examples of the differences with traditional standard map series. In map series, usually there is a standard scale with a standard sheet grid configuration. In producing census maps, the goal is to select a scale that works for a given operation that is centered on one geographic entity at a time. The entity may be a county or an enumeration assignment area. The sheet configuration also is unique to that entity.

Insets are automatically created when larger scales are needed for limited areas that noticeably reduce the total number of map sheets for a given entity. Insets are created in one of two ways: by rectangular areas within a parent sheet; or by clipping an irregularly shaped entity of dense content from within the parent map and insetting the amorphous area on the inset map.

Sheeting and insetting are iterative processes. Testing the conditions against results can lead to automatic changes in the parameters which initiate another round of testing within these CAMPS modules. When iterative processes occur, measures are invoked when preset options are exceeded. These limits stop the process and send that particular map entity into a referral system for a cartographer to review.

Symbolization associates a predefined map symbol with its feature. Symbol characteristics include: line type and weight, areal pattern, color, and specially designed point symbols. Information about how to place the symbol in the form of instructions is contained within the symbolization module. For example, many census geographic areas are hierarchical and form conjoint boundaries when they are coincident. It is not possible to symbolize multiple linear boundaries along the same string of coordinates and easily differentiate them. Intentional displacement of boundaries is not acceptable. In this case, alternating boundary symbols are displayed in a hierarchical order (Figure 4).

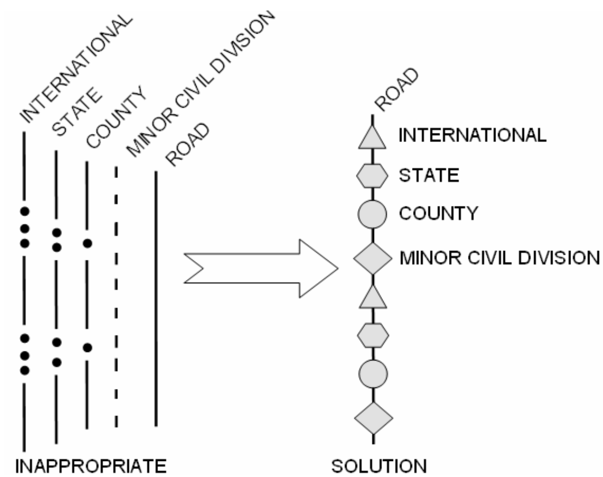

Figure 4. Alternating boundary symbol use instead of offsetting linear features.

Historically, text placement has been one of the most labor intensive and challenging activities when constructing a map. In interactive mapping, placing text has constraints that are little different from manual mapping where burnishing a name imprinted on a piece of thin plastic film to artwork with symbolized features first had to take into account the space available for the name. The process can involve multiple moves of names to place them. In some instances, constraints exist. For example text size is sometimes used as a means of representing data classes; altering the size to “fit” is not necessarily allowable (such as when text size is used to represent a particular data class).

Within CAMPS, like all of the other modules, text placement is automated. Most census maps require that all names are placed. In those instances, iterative processing continues to attempt the preferred placement. Like other modules, text placement is parameter driven. There is a point where labels are forced and that information is noted within CAMPS. In the limited instances where this occurs, the results are evaluated by cartographers to assure map readability.

The last phase of the mapping system centers on marginal information. Generally, content outside of the image area is either static or dynamic. Notes about the map series, or a standard legend, are examples of static information. Corner coordinates, the bar scale, barcode, and map sheet index and map locator are examples of dynamic data that change from sheet to sheet. Some elements contain both static and dynamic content where, for example, a map title identifies the map series as static information while the name of the specific mapped entity changes.

Quality Assurance and Quality Control

Quality control is an important component of any mapping process. In creating the tens of millions of maps for field operations, there are four stages of quality control to assure map use (Figure 5). A preproduction review occurs as the first phase where census cartographers review sets of maps for every field operation. Many of the maps selected represent geography that is difficult to map. During production, the second phase is a fully automated check based on selected rules and parameters within modules of the mapping system flow. For example, if the maximum allowed number of map sheets for a given entity is exceeded, an error message alerts a cartographer to review the results. This may lead to changes in the parameters.

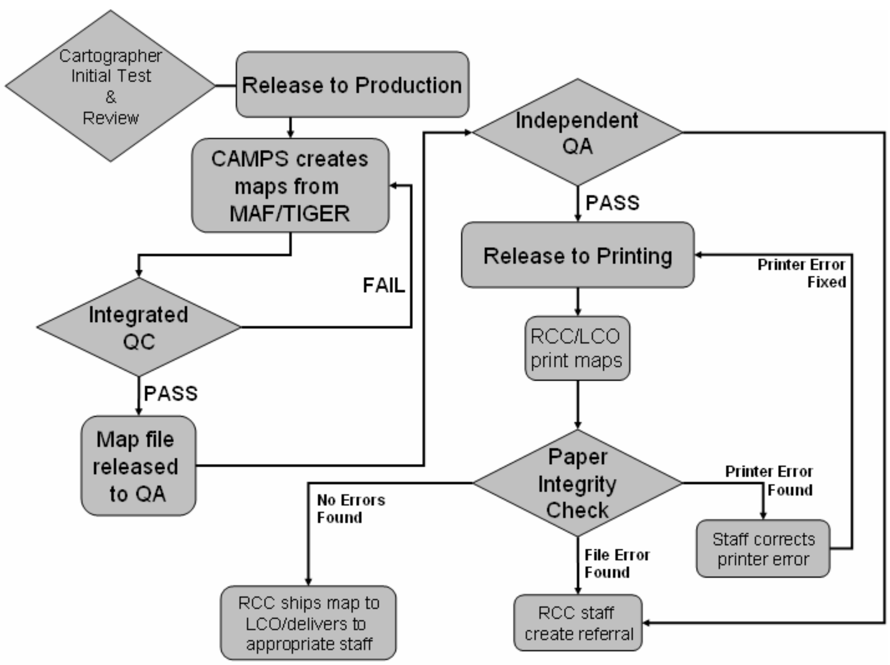

Figure 5. Quality assurance and quality control general process

During each map production operation, a sample set of map files are selected as part of the production run for the third quality phase. Normally, an initial sample of maps is reviewed followed by a controlled sample throughout map production. Geographers in Regional Census Centers (RCC) access a list of map files to review either on the screen or as paper maps. The list of map files is indicated within a separate map control system. While map files are in review, production continues but map files are not released to printing until the field staff approves the sample set. In this way, the workload continues which helps to assure that overall schedules are met. The last phase occurs at the printing stage where staff review the paper maps as they plot using large format plotters or on one of the two thousand small format printers in hundreds of decentralized locations across the nation.

Problems that occur are called referrals. When problems are identified, they are entered into a Problem Referral System (PRS). The PRS is a way to manage the status of specific problems. Entries in the PRS document the issue as it is passed from one person to another until the problem is resolved. Each referral is marked as fixed or accepted to assure that each entry is properly reviewed.

Based on automated notification, cartographers review the referral and classify its type for resolution into one of three processes based on its characteristics (Figure 6). Cartographers determine if the problem is a map specific issue, usually based on geographic anomalies. Parameters are adjusted to resolve the problem. In the second set of cases where a problem is systematic, cartographers evaluate the problem and implement corrective action. This step usually requires adjustment to mapping software or reconfiguration of parameter files. In the third scenario, a map file is corrupted. In this situation, the file is simply recreated.

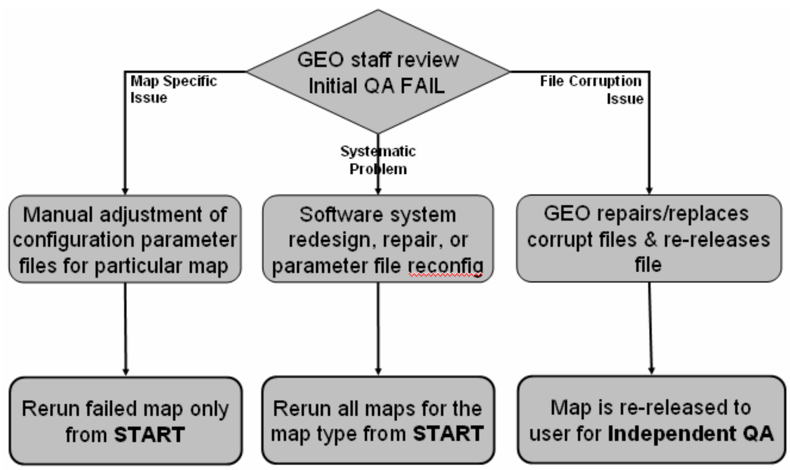

Figure 6. PRS resolution

Cartographers do not fix data errors. The goal is to identify the root cause of the error so that the correction is manifested throughout the geographic support system. If a data error is identified, that information also is entered into the PRS and directed to the appropriate staff for evaluation. If the error is critical, the data are fixed and the affected databases are rerun to assure that products contain the correct data. If the error is not deemed critical, the correction is made to the MTDB and reflected on future derivative databases and products.

Metadata

Map metadata is created during the mapping system flow and is used to manage production, dissemination, printing, MTDB updating, and problem resolution. Each map file has unique metadata. The map production workflow is dependent on information contained within metadata records. A variety of data, beyond the standard metadata records, is captured and used as part of each map production phase. The information is read by CAMPS and processed and is used by subsequent phases. As the map file builds, the metadata record is populated with additional information.

Upon completion and availability of metadata, other production processes use the information as part of their systems. The following examples emphasize its useful diversity. The availability of metadata is a signal to production that the map file is created which then populates production reports and is ready for dissemination. A map printing control system reads the metadata for determining the correct printer and order of printing. The MTDB update software reads map sheet corner point information to access the geographic window for updating and also projects the geospatial data based on the map projection. The PRS creates its reporting shell based on the universe of map files in case a problem is reported.

Metadata's use is not limited to current production. Information is used to project map sheet estimates for future mapping operations. Studying the results of past mapping provides important data about acceptable map scales. In the map scale example, mapping projects that have similar requirements may use metadata records of past operations to save on future processing-intensive activities, and giving relief to production schedules. The value of metadata is clear in its extensive use.

The Map Production Workflow Process

Organizationally, cartographers design maps and develop mapping systems. Once a mapping system is ready for production, it is deployed to a staff that performs independent testing and verification to assure that the production process will run successfully. Census operation schedules generally assume an equal flow. In a 24 x 7 production environment, each day's anticipated results are carefully calculated. There is very little lag in the schedule and there is no scheduled downtime while in production.

The Workflow Control Staff (WCS) manage production computer resources, stage the production runs, monitor the production process and disseminate production results. Activities that interact with and/or use the MTDB coordinate their work through the WCS. Mapping is an example of one of these activities.

A web-based control system (WebCS) manages all census geography batch processes. For example, creation of a MTDB benchmark, creation of derivative databases, geocoding operations, and batch feature updates to the MTDB are examples of fully automated processes. Once these processes are initiated, they continue running until completion, regardless of the time or day.

Initial job parameters are established. The criteria are programmed into the production system. Conditions for releasing to production include prerequisites such as the availability of a MTDB benchmark, a product database, and a cartographic database. The status of each dependency is automatically read from the WebCS Oracle tables. The order that determines which process to run is marked in a file that contains the universe of entities for production.

CAMPS is initiated by the WebCS which also uses Cronacle, a business process and scheduling software. WebCS executes CAMPS in which all of the mapping processes automatically run and are completed when a map file ultimately is created in PDF format. Each mapping run is processing intensive. The WCS controls the number of mapping runs submitted at any given time, depending on the number of available computer processors and disks assigned to the operation.

The mapping run invokes the initial job parameters (Figure 7). The map parameters from the Oracle tables are used along with MTDB system tables. As the mapping system programs run, sheet configuration determines the number of parent map sheets. Sheet finalization takes the sheet configuration results and prepares the map index to sheets. Geometry processing selects features, applies the map projection, and performs appropriate clipping against the image area extent.

Throughout the initial map file generation, information is passed to work tables that ultimately are used for creating the map file. After the initial map file is generated, text placement software (TPS) applies rules previously established in parameters based on the original map requirements. Adequate space for text placement is set as an earlier condition in the scaling algorithm. If text placement is unsuccessful, then the process reiterates and establishes new conditions. If conditions exceed acceptable limits, the process is stopped and passed to the PRS. This is a rare event due to the amount of testing and planning prior to production.

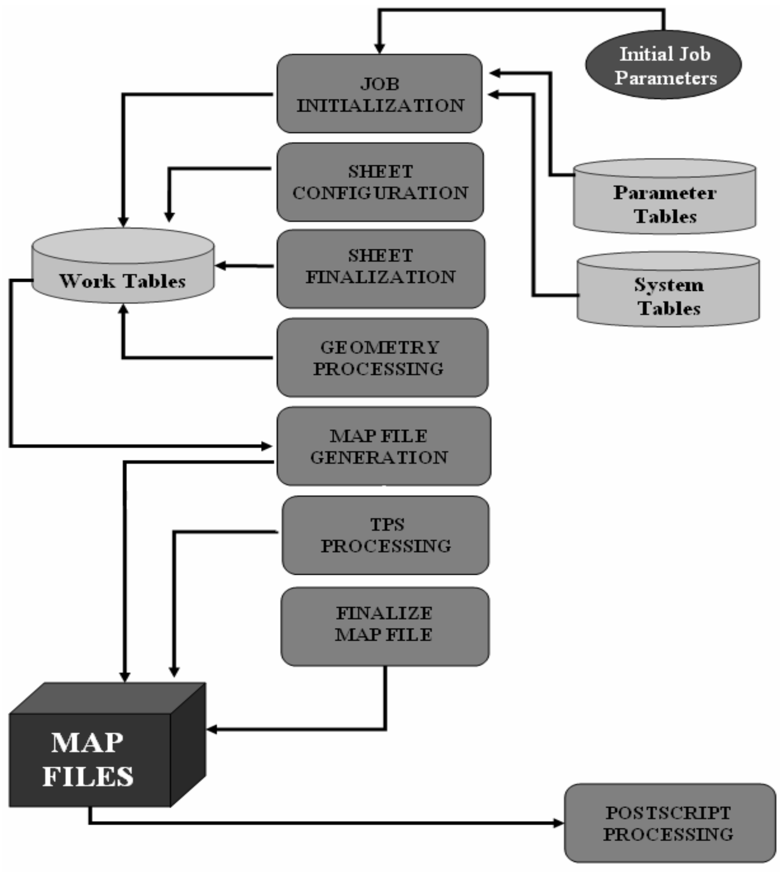

Figure 7. CAMPS Modules

The map file format is the final step in creating a map file. The CAMPS modules take place in an Oracle environment. Successful creation of a map file transitions to an ASCII-based format, which is converted to a Postscript format. Map files are then converted to PDF, the final map format. The files are stored on blade servers and await their final disposition. For field maps, the map files are either pushed to data processing centers for access by one of 494 temporary local census offices with small format HP 9050 printers or they are made available to one of 12 RCCs and a National Processing Center for plotting on large format HP 4500 color plotters.

Conclusion

The scope of mapping for a census is an enormous task. When interacting with sister statistical organizations around the world, the one activity that is challenging more than most is the effort required to prepare maps for census takers to adequately do their job. The sheer number of maps needed to cover a nation at workable, usable scales is quite a task. Complicating that further is the daunting schedule to meet map user needs for large field operations.

The size of the United States precludes manual approaches for making maps for a census. Computers running well developed fully automated mapping systems are the only choice for meeting the mandate. Highly skilled cartographers ingeniously planned, developed, and implemented the Census Automated Mapping System to meet the demanding requirements.

Data and software alone are not a panacea to successful census map support. A workflow process that ensures sufficient time and resources to meet a production goal are critical to success. The U.S. Census Bureau has combined geospatial data bases, sophisticated mapping software systems and carefully crafted production workflows to assure a successful outcome.