Section 5 – Summary of Methods and Processes

This section summarizes the various innovations, methods and processes created or used in the completion of this project.

INNOVATIONS

Geocoding Approach

One of the most rewarding parts of this project was the development of the enhanced geocoding methodology described in detail in Section 6. This enabled addresses with ZIP code information only to be geocoded to census block group for the purposes of mapping. This innovative method virtually eliminated concerns about bias in the data represented, which was once a concern for rural poverty areas, and resulted in overall geocoding success rates at the census block group in excess of 96% for the combined grantee's data.

GeoStatistical Abstract

While not a mapping product, another new product created by this project was the "geostatistical abstract". Summaries of available census poverty data along with tabular statistics regarding the grantee case data and the creation of simple charts were of great interest to the grantees and of assistance in the design of meaningful maps. These statistics include grantee service area poverty demographic data, resulting in an overall statistical view of the data most relevant to the grantees. A copy of the geostatistical abstract is included in Appendix A.

Mapping Software

For this mapping evaluation, the latest version of ESRI's ArcView software, ArcGIS 8.3 was used. For the previous Mapping Evaluation Legal Services in Georgia, ArcView 3.2 was used to create the data and maps. The benefits of ArcGIS 8.3 include improved data management and flexibility, as well as enhanced cartographic abilities such as halo text options.

Geodatabase

The geographic base data for this project were stored in geodatabase format, which utilizes Microsoft's Access database. Combined with the data stored in SQL Server and the live connection capability of the layer file format, access to map data was fully integrated. It also made the data more reusable for mapping. Many redundant steps necessary to establish symbolization across multiple maps was automated using layer files and geodatabases.

SQL Server

The Census demographic and grantee case data were stored in Microsoft's SQL Server. SQL views were created to select the specific case or demographic data for each map and perform the required ratio or density calculations. The mapping software connects to these map specific SQL views and joins this virtual table to the appropriate geography. This connection between geography and SQL view is stored in a 'layer file' that also stores the symbology and legend.

Live Connection with Layer Files

The newest version of mapping software introduced a new file format, layer files, that took the place of the traditional shape files for use in connecting directly to the map data contained in the SQL Database. The “live connection” that was created allowed for automatic map updates whenever the data was updated. This was far more efficient than the shape file updates used in the initial project in Georgia.

Halo Text

Another new feature in the mapping software was Halo Text. By placing a white halo around text (specifically office labels), those labels considered of high significance can be made to “pop” off the map. Due to the large amount of data displayed on the maps, the halo text had a positive impact on the overall cartographic product.

Map Sheets

Map Specification sheets were produced to assist the cartographers in the mapping process, the audience with map interpretation and with quality control. New to the Georgia project was the use of Microsoft Access in the production of the map sheets. All information pertinent to the maps was maintained in Microsoft Access. A sample of the map specification sheets produced can be found in Appendix A.

Legends

A separate mapping document was created containing all of the legends required for all maps created during the project. Because the legend titles were not standardized, the legends could not be automatically generated or updated. In order to maintain consistency and efficiency, a single document with all legends was created that allowed the cartographer to cut and paste the appropriate legend into the appropriate map document with out having to generate a “new” legend for each map.

Reference Scale

The ability to efficiently produce zoom in maps was greatly enhanced due to the addition of the Reference Scale feature in the mapping software. By changing the reference scale on the map, the computer automatically adjusted the font size and line widths to correspond with the scale of the zoom in map. While beneficial, this feature is only highly effective if Dynamic Labeling is utilized.

Zoom In Maps

A stated goal of this project was to investigate the opportunities for more neighborhood scale or zoom in maps. This goal was accomplished through a combination of street address level geocoding and the production of very high quality digital PDF files. The users are able to digitally zoom in on areas of interest to visualize greater detail or print hard copy at any larger sheet size as needed to see the additional detail. This ability supported including the entire service area for any map data set produced, since it was useful to have the entire service area available for this digital zoom in capability.

Matters Mapping

This project included an attempt at mapping of matters data. While much of this data had insufficient address information to be mapped, some of these maps were produced. In particular, a "Total Activities" map that displays a distribution of both case and matters data over poverty population density seems a likely candidate for some future mapping projects.

GEOGRAPHIC DATA

In order to ensure uniform comparisons the vintage and source of all geographic data to be used should be established early in the project. This is even more critical in the case of a nationwide program where data may be compared on multi-state or national levels in addition to the local, state and regional perspectives used in this project. The purchase of geographic data for the entire U.S. versus by individual state offers opportunity for reduced costs in a nationwide program.

An issue that could have an effect on a nationwide mapping program is the challenge of establishing coordinate systems for the maps. On this project, the universal geographic coordinate system was to be used for all data and maps were to be projected "on the fly" into the local state plane coordinate system.

However, during actual execution of the project, the Census and base data did not project on the fly correctly. While projecting on the fly was acceptable for display purposes, accurate analysis and selection across themes was not possible. The solution to this issue was to pre-process all of the geographic data and project it into the state plane coordinate system before storing it for mapping use. This approach would be a significant effort for a nationwide mapping program that would require projection of numerous data sets into the state plane coordinate systems of each region of each state.

Country, State, County and Census Geography

Country, state, county and census geography used for this project was 2000 Census TIGER files enhanced by Tele Atlas (GDT, Inc.). This same geography was used for both geocoding and mapping to provide consistent data results between these two separate processes.

A technical issue with the specific geography used on this project involves the treatment of water features such as oceans, lakes and rivers in the data. Water features were not included in the geography files. The result of this is "holes" or "blank areas" in the data wherever water features were identified. This created several cartographic challenges.

First, in the ArcGIS maps, these holes are transparent through to the background color of the display or the map. This may be unsettling cartographically. The approach used to overcome this was to set background color to a blue color that represents water features.

The disadvantage of this solution is that background colors must be set to the intended water feature color and cannot be used to enhance the map in other ways. In addition, this requires active manual action and deviates somewhat from the intended automation objectives. For a national program, it may be necessary to ignore the cartographic color of these water bodies or to uniformly specify the color to be used for the background to represent water features.

Second, because they are not included as part of the geography, the water features have cartographic boundary lines around them to indicate the boundary of the geography. This created visual confusion regarding state and county boundaries. The solution selected to correct this was to hand edit the water features to remove the cartographic boundary lines.

The disadvantage of this solution is the labor-intensive effort required to change these boundaries and the deviation from the intended automation objectives of the project. For a national program, it would likely be resource prohibitive to address this cartographic boundary concern.

Lastly, the detail of water boundaries along the coastlines, rivers and other major bodies of water creates a cartographic issue when data from other sources overlays these areas. The boundary is so detailed that it cannot be displayed clearly at the map scale being used for this project. The display appears to have additional lines along these boundaries. This cartographic issue was corrected by systematically and manually "smoothing" the boundaries so that they were acceptable.

As with the boundaries in general, the disadvantage of this solution is the labor intensive effort required to modify these boundaries and the deviation from the intended automation objectives of the project.

USPS ZIP Geography

USPS ZIP geography used for this project was USPS data enhanced by Tele Atlas (GDT, Inc.). This same geography was used for both geocoding and mapping to provide consistent data results between these two separate processes.

Street Geometry

Street geometry used for this project was 2000 Census TIGER files enhanced by Tele Atlas (GDT, Inc.). This geometry was used for the geocoding process only. For mapping purposes, the base layer geometry described below was used. This technique did not provide consistent data results between the two separate processes, but was chosen to simplify the cartographic mapping effort.

Base Layer Geometry

Interstate, road, water feature, and city point geometry used for mapping on this project was Environmental Systems Research Institute (ESRI) data. This geometry was used for the map reference and annotations purposes only. For geocoding purposes, the GDT street geometry described above was used. This did not provide consistent data results between these two separate processes, but was chosen to simplify the cartographic mapping effort.

In order to facilitate automated mapping processes, specifications for data to be included as base layers needs to be identified early in these projects. These base layer specifications include specific information regarding what base layers appear at various scales and the criteria for inclusion of specific features. A possible alternative to using the geography displayed on the map to assign base layer criteria may be the selection of the map as being urban or rural. However, the vast difference in geometric size of states and counties may make this impractical. Another alternative might be to base these criteria on actual map scale, independent of the geography displayed on the map in order to mitigate the diversity in state and county geographic size.

Transportation, Water and City Features on Maps

The type and intended use of maps determined the way base or reference features are displayed on a particular map. Map makers must decide on the display of roadways with separate criteria for interstates, freeways, other federal highways, state highways, county roads, local roads, etc., and must decide the cartographic display characteristics as well.
Similar decisions arise for water features with criteria for oceans, lakes, ponds, rivers, and streams. The criteria for display of cities on state level maps might include those with population greater than 500,000, whereas cities with populations greater than 25,000 may be more appropriate for county level maps.

A technical issue with the city geometry used on this project involves the location of the center of the city. The data used for mapping displayed the geometric center of the city as a point. This created a cartographic challenge, since the perceived location of cities, particularly those along coastlines, does not coincide with the geometric center of the actual city boundary.

The approach used to overcome this was to manually locate the city points to closely represent the perceived location of the cities. The disadvantage of this solution is that it requires active manual intervention and deviates from the intended automation objectives. For a national program, it may be necessary to either ignore the cartographic issue of location for these cities or investigate additional data sources that may provide more pleasing geometric location data for cities.

An additional cartographic issue associated with the city geometry on this project was dissatisfaction with the criteria used for the cities to be displayed. In this case, the number of cities shown using population based criteria was deemed unacceptable cartographically. The approach used to overcome this was to request custom lists of cities to be displayed for each map template and to manually locate the city points representing the location of the cities.

The disadvantage of this solution is that it requires custom lists to be generated for each different template as well as active manual intervention, which both deviate from the intended automation objectives. For a national program, it may be necessary to either ignore the cartographic issue of the number of cities shown or specify definitive criteria that will be acceptable.

Supplemental Geometry

Office Location Geometry

Grantee Office point geometry used for mapping on this project was developed by obtaining office address information from the Grantees and geocoding their locations from this address data.

The Grantees have a main office and some additional offices or facilities. The total number and location of these offices and facilities change from time to time. In order for the maps to display relative data, it was determined that the office and facility locations shown should be those that existed during the same time period as the case data being represented. In addition, we standardized on three types of offices to be displayed: Main Office, Branch Office and Sub-Grantee Office.

Establishing the name to be displayed on the maps for each office, the type of office, the Office Code and the correct address for each office during the initiation of the project is critical to efficient map production.

As with base layer data, in order to facilitate automated mapping processes, the data to be included in the office location layers should be identified early in these projects. The office location layer specifications should include specific information regarding what offices appear at various scales and the criteria for inclusion, etc.

It was believed at the inception of this project that maps aggregated to the service area associated with a specific office might be useful. Instead, zoom in maps of specific areas of interest with aggregations at Census Tract or Block Group geographies were deemed more informative and useful. As a result, the location of offices and office service area boundaries were used for visualization and orientation purposes only.

Service Area Geometry

Grantee service area geometry used for mapping on this project was developed by obtaining a listing of Census or USPS ZIP geographies from LSC that make up each of the Grantee service areas. In addition, similar data was collected from the Grantees for the office service area boundaries. These geographic areas were then grouped and boundary lines were generated around them using the mapping software to produce the desired service area geometry.

The number of Grantees and their service areas are fairly stable, but they may be revised from time to time. In addition, the Grantees have a main office and some additional offices or facilities. The total number and location of these offices and facilities change from time to time. Accordingly, the service areas associated with the Grantees and their offices may change.

In order for the maps to display relative data, it was determined that the service area geometries shown should be those that existed during the same time period as the case data being represented.

As with base layer data, in order to facilitate automated mapping processes, the data to be included in the service area layers should be identified early in these projects. The service area layer specifications should include specific information regarding what service area boundaries appear at various scales and the criteria for inclusion, etc.

It was believed at the inception of the project that maps aggregated to the service area associated with a specific office might be useful. Instead, zoom in maps of specific areas of interest with aggregations at Census Tract or Block Group geographies were deemed more informative and useful. As a result, the location of offices and office service area boundaries were used for visualization and orientation purposes only.

CSR Case, Non-CSR Case and Matters Data

CSR Case, Non-CSR Case and Matters Data geometry used for mapping on this project was developed by obtaining complete case record address information including case record number, street address, city, state and ZIP code from the Grantees and geocoding their locations from this address data.

The geocoding process and statistics regarding the results can be found in Section 6 of this report.

Deemed Duplicate Records

Upon inspection, the case data submitted by Grantees contained records that appeared to be duplicates.

The determination was made that all records that met the following seven criteria be flagged as deemed duplicates, and submitted to the Grantees for review:

The Client Name is the same AND
The Client Address is the same AND
The Client Gender is the same AND
The Client Ethnicity is the same AND
The Client Age is less than 3 years apart AND
The Problem Code is the same AND
The Case Closed Date is in the same year.

Following the review, any case records that were incorrectly flagged as deemed duplicates were left in the data set. All others were removed from further processing and were not mapped. This process resulted in the removal of 6,794 records, approximately 5% of the total records submitted.

TABULAR DATA

CSR Case Data

Legal Problem Category and Legal Problem Code

There are ten Legal Problem Categories with each broken down into several Legal Problem Codes as shown in Appendix A.

Data submitted by the Grantees may contain Legal Problem Category and Legal Problem Code information. This allows the data to be mapped with a focus on a particular Legal Problem Category or Legal Problem Code.

The actual data submitted by Grantees sometimes included Legal Problem Codes and descriptions that did not conform to the standard Legal Problem Codes provided by LSC. In addition, some Grantee data was broken down to sub-Legal Problem Codes with their own descriptions.

In general, having numerous ways of reporting this data created several data management challenges. When only Legal Problem Codes were provided, these Legal Problem Codes had to be mapped to Legal Problem Categories, an extra processing step that must be customized for the specific data. Further, when non-standard Legal Problem Codes are provided, no basis for mapping them to Legal Problem Categories exists, requiring that these non-standard Legal Problem Codes first be assigned to a standard Legal Problem Code before mapping them to Legal Problem Categories. For this project only a small number of maps were developed by Legal Problem Category, with none produced using the Legal Problem Code data. However, in order to maximize the opportunity for mapping using these data sets in the future, it may be useful to continue including both Legal Problem Category and Legal Problem Code as a part of the standard data set.

For a nationwide mapping program, it will be imperative that the data be standardized to include both Legal Problem Category and Legal Problem Code data (if desired) that fully conform to the old or a new more complete standard established by LSC. The practice of providing only Legal Problem Codes further complicates the data processing as discussed above.

CSR Case Closure Category

There are eleven CSR Case Closure Categories as shown in Appendix A.

Data submitted by the Grantees may contain CSR Case Closure Category information. This allows the data to be mapped with a focus on a particular CSR Case Closure Category.

The actual data submitted by Grantees sometimes included CSR Case Closure Categories and descriptions that did not conform to the standard CSR Case Closure Categories provided by LSC. When non-standard CSR Case Closure Categories are provided, no basis for mapping them exists, requiring that these non-standard CSR Case Closure Categories first be assigned to a standard CSR Case Closure Category.

For this project, the CSR Case Closure Categories were further grouped into four “Closure Mapping Groups” as shown in Table 5.1.

Table 5.1

CSR Case Closure Category	Closure Mapping Group
A, B	Brief Service
F, G, H, I, K	Extended Service
C	Referral
D, E, J	Other

Only a small number of maps were developed by Closure Mapping Group, with none produced using the individual CSR Case Closure Category data. However, in order to maximize the opportunity for mapping using these data sets in the future, it may be useful to continue including CSR Case Closure Category as a part of the standard data set.

For a nationwide program, it will be imperative that the data be standardized to include CSR Case Closure Category data that fully conform to the standard established by LSC.

Office Code

Data submitted by the Grantees may contain Office Code information to reflect the office providing service for each case. This allows the data to be mapped with a focus on a particular Office.

The actual data submitted by Grantees sometimes included Office Codes and descriptions that did not conform to the list of offices provided for Office Location Geometry (discussed previously).

When non-standard Office Codes are provided, no basis for mapping them exists, requiring that these non-standard Office Codes first be assigned to a standard Office Code for which Office Location Geometry has been provided.

It was believed at the inception of this project that maps aggregated to the service area associated with a specific office and including case data based upon the Office Code might be useful. Instead, zoom in maps of specific areas of interest with total data aggregations at Census Tract or Block Group geographies were deemed more informative and useful.

As a result, the location of offices and office service area boundaries were used for visualization and orientation purposes only, and no specific data based upon Office Code were mapped. However, in order to maximize the opportunity for mapping using these data sets in the future, it may be useful to continue including Office Code as a part of the standard data set. Further, establishing the name to be displayed on the maps for each office, the type of office, the Office Code and the correct address for each office during the initiation of the project is critical to efficient map production.

For a nationwide program, it will be imperative that the data be standardized to include Office Code data that fully coincide with Office Location Geometry.

Supplementary Case Data

Data submitted by the Grantees may contain Private Attorney Involvement, Funding Source or Code, Outreach Sites, Special Programs, Clinics, Language, Family Size, Family Type, Income Source or Code, Residence Type, Education, Gender, Ethnicity, Intake Method or Code, Age, Case Opened Date and Case Closed Date information. This allows the data to be mapped with a focus on these particular criteria.

While some Grantees provided this data, it varied in format and categorization significantly. In general, having numerous ways of reporting this data created several data management challenges resulting in an extra processing step that must be customized for the specific data. For this project, only a small number of maps were developed using these data. However, in order to maximize the opportunity for mapping using these data sets in the future, it may be useful to continue including them as a part of the standard data set.

For a nationwide mapping program, it will be imperative that the data be standardized to include Private Attorney Involvement, Funding Source or Code, Outreach Sites, Special Programs, Clinics, Language, Family Size, Family Type, Income Source or Code, Residence Type, Education, Gender, Ethnicity, Intake Method or Code, Age, Case Opened Date and Case Closed Date data that fully conform to a standard established by LSC.

Non-CSR Case and Matters Data

The tabular data associated with Non-CSR Case and Matters data varied widely. There are no standards for the data to be included with these records requiring data management to be customized for each specific data set.

For this project, only a small number of maps were developed using Non-CSR Case and Matters data. However, in order to maximize the opportunity for mapping using these data sets in the future, it may be useful to continue including Non-CSR Case and Matters information as a part of the standard data set.

For a nationwide program, it may be useful to include Non-CSR Case and Matters data that fully conform to a new standard for data collection and reporting.

Demographic Data

Demographic data was extracted from tables provided by the U.S. Census Bureau.

Statistics including race and age were available only for 100% poverty levels restricting the use of comparative maps regarding race or age to that poverty level.

A summary of the census tables used is shown in Appendix C.

Supplemental Data

Various supplemental data was used for development of some Special Maps. This data all presented data management challenges as it was associated with differing geographies, with different source years, etc. In addition, there was often little documentation regarding the source or methodologies used to develop the original data.

While a few maps were produced utilizing this type of data, one map type became of particular interest and its production has been recommended for future projects. This map type displays the density of attorneys based upon data received from the appropriate Bar Association. The largest challenge with this particular data set was that the location of attorneys is based upon their ZIP code. The nature of ZIP codes is that they change in number, geographic location, etc. making it difficult to correlate the data provided to the ZIP geography being used for mapping. Fortunately, the geocoding software being used for this project contained conversion tables that could associate older ZIP codes with the current ZIP codes being used for mapping. This required an extra processing step, but facilitated the production of these Special Maps.

MAP TEMPLATES

Map Titles

The map titles for the standard maps on this project were established as a part of the scope definition and were included in the contract agreement.

On the original Georgia project there were five major map categories tested. Only two of these major categories survived in this project: choropleth and dot density. During the course of this project, it was observed that there were generally ten specific types of maps being produced. A map title standard was produced for each of these types and was used to generate revised titles for the standard maps as well as new titles for special maps.

A description of the ten basic types of maps produced is shown in Table 5.2.

Table 5.2

Map Type	Type Description	Graphic Description
Distribution	Dot location of data	Dots
Simple	Single data by aggregation	Choropleth
Percentage	Percentage of data by aggregation	Choropleth
Change	Absolute change in data by aggregation	Choropleth
Percent Change	Percent change in data by aggregation	Choropleth
Dot Density	Dot density of data by aggregation	Dots
Density	Density of data per square mile by aggregation	Choropleth
Ratio	Ratio of dataA to dataB by aggregation	Choropleth
Distribution over Density	Dot location of dataA over density of dataB per square mile by aggregation	Dots over Choropleth
Distribution over Ratio	Dot location of dataA over ratio of dataA to dataB by aggregation	Dots over Choropleth

Map Legend Titles

The map legend titles for the standard maps on this project were established as a part of the scope definition and were included in the contract agreement.

A map legend title standard was produced for each of the ten types of maps and was used to generate revised legend titles for the standard maps as well as new titles for special maps.

Generalized guidance used to develop map and legend titles for each type of map along with the complete legend key information may be found in Appendix A.

For a national program, it will be beneficial to establish the map title standard for each map type in order to ensure both consistency for end users and efficiency in the map production.

Map Legends

The map legends for the standard maps on this project were established as a part of the scope definition and were included in the contract agreement.

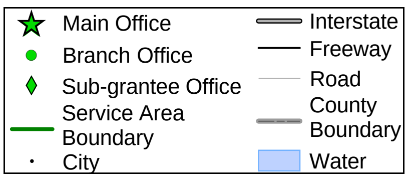
One of the goals of the project was to test the legends developed for the previous mapping project. More specifically, the legend classifications and breaks were to be tested for validity and usefulness in the highly urban areas of southern California.

A summary of the general legend classifications and breaks used for this project compared to those proposed from the previous project can be found in Appendix A.

Map Office Key

In this project, we standardized on three types of offices to be displayed, each with their own map symbol. These map types and symbols along with other base data symbols used for the templates are shown in Table 5.3.

Table 5.3



Map Notes

While an attempt was made to standardize on the map notes from the previous project, some modifications were made to incorporate comment provided by the Grantees. The standardization of these notes will be important to the efficiency of any nationwide mapping project.

Map Labels

The cartographic aspect of labeling Countries States, Counties, Cities, Offices, Transportation features and Water features is significant.

Initially, a technique known as dynamic labeling was to be employed for all of the maps with the benefit of being able to produce zoom in maps at different scales and having the labels adjust in size and orientation automatically (dynamically).

Through the course of the project, the compromises required for the use of this labeling technique proved to be unacceptable to the Grantees. This was resolved by hand placing labels for each Grantee map template and for each map of a differing scale.

During this process, a labeling technique called halo text was introduced and became useful in the production of cartographic labeling. This technique produces a white space outline or halo around text, setting it off from the surrounding data and making it more legible.

For a nationwide mapping effort, it is likely that cartographic concessions will have to be made to allow the use of dynamic labeling in the interest of both speed and cost. Enhancements to the mapping software may include additional labeling controls that would further minimize this impact for future projects.