



U.S. Department  
of Transportation  
National Highway  
Traffic Safety  
Administration



DOT HS 809 201  
NHTSA Report

April 2001

# Geographic Information Systems Using CODES Linked Data

(CRASH OUTCOME DATA EVALUATION SYSTEM)

## Technical Report Documentation Page

1. Report No.  DOT HS 809 201	2. Government Accession No.	3. Recipients's Catalog No.	
4. Title and Subtitle  Geographic Information Systems Using CODES Linked Data (Crash Outcome Data Evaluation System)		5. Report Date  April 2001	
		6. Performing Organization Code	
7. Author(s) (See Attached)		8. Performing Organization Report No.	
9. Performing Organization Name and Address		10. Work Unit No. (TRAIS)n code	
		11. Contract of Grant No. (See attached)	
12. Sponsoring Agency Name and Address  National Highway Traffic Safety Administration 400 Seventh Street, S.W. Washington, DC 20590		13. Type of Report and Period Covered NHTSA REPORT	
		14. Sponsoring Agency Code	
15. Supplementary Notes Dennis Utter served as the NHTSA Contracting Officer's Technical Representative for the study			
16. Abstract  This report presents information about geographic information systems (GIS) and CODES linked data. Section one provides an overview of a GIS and the benefits of linking to CODES. Section two outlines the basic issues relative to the types of map data, resources needed for developing a traffic safety GIS, institutional issues and the linking of CODES to GIS. It also includes implementation recommendations. Section three includes descriptions of the CODES/GIS developed by Hawaii, New Hampshire, Maryland, and South Carolina. The purpose of this information is to encourage other CODES states to expand CODES to include GIS.			
17. Key Words  Geographic Information Systems, CODES, linked data, probabilistic linkage		18. Distribution Statement Documents available to the public through the National Technical Information Service, Springfield, VA 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (Of this page) Unclassified	21. No of Pages 40	22. Price

## LIST OF AUTHORS

Preparation of this document was a collaborative effort involving persons from four CODES states, NAGHSR and NHTSA as listed below:

Hawaii	Karl Kim, University of Hawaii, Department of Urban & Regional Planning (DTNH22-96-H-39017)
Maryland	Tim Kerns University of Maryland National Study Center for Trauma/EMS (DTNH22-00-H-07012)
New Hampshire	Tom Hettinger New Hampshire Department of Safety, Division of Motor Vehicles (DTNH22-00-H-37012)
South Carolina	Mary Pease South Carolina State Budget and Control Board (DTNH22-00-H-17012)
Sandra Johnson, editor Consultant, National Association of Governors' Highway Safety Representatives (DTNH22-94-H07364)	
Dennis Utter, CODES Contracting Officer's Technical Representative National Highway Traffic Safety Administration	

## TABLE OF CONTENTS

OVERVIEW OF GIS AND CODES - Karl Kim, Hawaii CODES .....	1
--	---

### IMPLEMENTATION ISSUES FOR CODES/GIS

GIS and Types of Map Data - Karl Kim, Hawaii CODES .....	9
--	---

Development of a Traffic Safety GIS - Mary Pease, SC CODES .....	13
--	----

! Recommendations .....	16
-------------------------	----

Institutional Issues Tim Kerns, Maryland CODES .....	17
--	----

! Recommendations .....	18
-------------------------	----

Linking CODES to GIS - Tom Hettinger, NH CODES .....	19
--	----

! Recommendations .....	20
-------------------------	----

Analytical Issues Karl Kim, Hawaii CODES .....	21
--	----

! Recommendations .....	23
-------------------------	----

### STATE SPECIFIC DESCRIPTIONS GEOGRAPHIC INFORMATION SYSTEMS

Introduction .....	25
--------------------	----

Hawaii GIS System .....	26
-------------------------	----

Maryland GIS System .....	30
---------------------------	----

New Hampshire Proposed GIS System .....	33
---	----

South Carolina GIS System .....	37
---------------------------------	----

## OVERVIEW OF GIS AND CODES

GIS (Geographic Information System) combines computer hardware, software, databases and personnel to manage, display, map, and analyze information related to spatial phenomena (Figure 1). Advances in desktop mapping, GIS software packages, integration between GIS and GPS (global positioning systems) as well as general improvements in computer systems have helped to provide new tools for traffic safety research. These tools are becoming increasingly powerful and useful for problem definition, program design and evaluation, and various educational, enforcement, and engineering applications designed to enhance highway and traffic safety.

The purpose of this report is to summarize information about GIS implemented by the CODES (Crash Outcome Data Evaluation System) states. CODES staff from Hawaii, New Hampshire, Maryland and South Carolina collaborated with each other to provide the information about GIS they considered most important for CODES states interested in developing GIS.

The CODES project, funded by the U.S. Department of Transportation, National Highway Traffic Safety Administration, is a data-driven project. CODES involves the linkage and analysis of police crash data, emergency medical services transport data, hospital data, and insurance claims. CODES data often contain location information such as County, Census Tract, Police Beat, street or roadway name, street address, distance from a milepost or nearest intersection or references to other geographic features and place names. Several of the states have linked their databases to base maps to permit analysis of spatial trends and patterns of motor vehicle crashes and the resulting injuries.

This report consists of three sections. Section one provides an overview of a GIS and the benefits of linking CODES and GIS. Section two presents information about the basic issues relative to GIS. This information was originally presented to all of the CODES states and other states interested in becoming CODES states at a technical assistance meeting, CODES TA 2000, in Portland, Maine June 19-21, 2000. The issues discussed addressed different types of map data, resources needed for developing a traffic safety GIS, institutional issues and the linking of

Figure 1

CODES to GIS. Recommendations for successful implementation relative to each group of issues were developed by the four CODES states and presented for discussion by the group as a whole. The final versions of the recommendations also are included in this section. Section three presents the state specific descriptions of the CODES/GIS developed by Hawaii, New Hampshire, Maryland, and South Carolina.

The nexus between CODES and GIS is strengthened by several different factors. First, many of those within the collaborating agencies, police, EMS, public health specialists, highway engineers, and others, are interested users of spatial data (maps, GIS and GPS). Second, many of the strategies for problem definition, program design and evaluation of traffic safety countermeasures rely explicitly on location information. Questions are often posed in terms of point locations (which intersections produce the highest number of fatal crashes); or segment and roadway queries (which roads have the highest incidence of bicycle or pedestrian crashes); or zonal or areal tabulations and analyses (how do cities or towns or census tracts or block groups compare in terms of the frequencies of various types of crashes). With linked data, however, the questions and topics of inquiry can be expanded to include more detailed information about ambulance transports, medical treatments, hospital and insurance costs, and other elements contained within the CODES linked databases. Which highways produce the most EMS runs? Which segments of roadway have the highest costs in terms of utility pole crashes? Which produces more hospitalizations among elderly drivers involved in crashes at a particular location - broadside crashes or head on collisions? GIS is an enabling technology that allows safety researchers to fully explore relationships between crash and injury variables contained within the CODES databases.

In recent years, the GIS software market has been dominated by: ESRI's Arc/Info and ArcView; Integraph's GeoMedia; MapInfo; and Caliper Corporation's TRANSCAD software. There are others that are available, but they are not widely used. Like the dominance of the Statistical Analysis System (SAS) within the federal government and in many state government agencies, these software packages have grown to be widely taught at universities and implemented in many different organizations. In addition to the increased use of menu-driven command languages, the software has also been enhanced through the development of many different subroutines, extensions, and other add-on software packages that add a variety of different functions to the core software package. Some of these enhancements allow for a three-dimensional analysis of effects; others enable users to tabulate and report on various spatial statistics; still others provide algorithms for network analysis and other more specialized functions.

There are different implementation approaches each requiring different hardware configurations for developing GIS. Since many of the CODES sites already have computer equipment, GIS could be thought of as an add-on to existing systems. Many GIS packages such as ArcView or GeoMedia or earlier desktop mapping programs are available for Windows, Windows-NT, and Unix systems. Depending on the size of files

and on the requisites in terms of processing and display speeds, adding a GIS package to some systems may serve its purposes just fine. There may be a need to add additional hardware such as plotters, perhaps additional memory or secondary storage, since many contemporary GIS packages are both memory and computationally intensive. The approach here, however, is to build GIS on existing hardware, perhaps on the same platform where the principal CODES databases are stored and analyzed.

Another approach is to build a separate GIS system, using more customized hardware and software. Increased processing power and speed, as well as additional mass storage may be needed because many of the spatial databases can be very large. This may be particularly true if additional base maps (TIGER, digital line graph (DLG), digital elevation model (DEM), cadastral, etc.) as well as aerial photographs, satellite imagery, and other databases are integrated into the system.

Some DOTs have legacy systems integrating autocad functions, highway design and engineering, mapping, construction management, and GIS functionality. Rather than build a separate GIS system with CODES data, it might be more appropriate to add CODES data elements to these existing systems. In other cases, CODES data could be integrated with HPMS (Highway Performance Monitoring Systems), video logging systems, and real time highway information systems for congestion management, incident detection, or freeway management.

GIS can be thought of as another type of database management system. Most GIS packages recognize three different types of geographic features: points, lines, and polygons. Attribute data are stored or "linked" to these geographic features. In addition to attribute data, a GIS needs to have a base map and a reference dictionary describing the various geographic features contained on the base map. The process of geocoding involves assignment of geographic coordinates to attribute data so that they can be located on a map. Typically, latitude and longitude are used, but other coordinate systems may also be employed. Many different geographic products are used to map geocoded data. Most of the products vary in map scale and accuracy, thereby offering different advantages to a user. The US Bureau of the Census TIGER files are created at a 1:100,000 scale, and, although updated annually, vary in accuracy due to differences in their updated sources. Other common map products utilized are the US Geological Survey (USGS) digital line graphs (DLG) and digital elevation models (DEM) that range in scale from 1:24,000 to 1:100,000. These sources are periodically updated based on USGS needs and other criteria. All USGS files adhere to National Map Accuracy standards that meet a consistent scale. Additional map sources are cadastral or parcel layers usually based on aerial photography. These data sources are usually created at a 1:500 scale or larger. In addition to the map databases, it might be useful to include aerial photographs, satellite imagery, and other databases that can be used to identify and analyze locations. The declassification of spy satellite technology has meant that for example, the IKONOS satellite can produce stunning panchromatic imagery with a one-meter resolution or multi



spectral images with a four-meter resolution.

Not all of the CODES states have used the same base maps when developing traffic safety-related GIS. The most common base maps are the U.S. Bureau of Census TIGER files (Figure 2). They are widely used for several different reasons. First, they are comprehensive. Virtually, every part of the U.S. has been mapped. The TIGER files were developed to deploy Census workers so they have detailed information including address ranges for streets. Second, the TIGER files provide a uniform system for mapping and adhere to Federal Geographic Data Committee (FGDC) Metadata and content standards. The uniform system involves including enumerator updates, based on field drawings, that are incorporated into the digital files through heads-up digitizing. In other words, some street segments are added to the TIGER files based on field sketches from enumerators and then manually digitized while eyeballing these sketches. This technique, thus, is not very accurate or consistent in quality since each person sees the world differently. There are much technical support and documentation on using Census files. Because they are used in every state, third party vendors have updated the Census files and increased the functionality of the Census TIGER products. Finally, using the TIGER files as the base maps enable the integration of other Census products and data sources. Information about population characteristics, household size, employment, travel behavior, and other socioeconomic factors can be easily brought into the analysis.

Figure two, "TIGER Line Files - Road Network," shows an example of a base map and attribute data available in the Census TIGER files. This figure shows the road network base map for the island of Oahu, Hawaii, as well as a few of the common attributes available within the TIGER system. Within the TIGER files, there are a number of attributes associated with each street segment, but the most widely used for the geocoding process are:

- Fname - Street Name
- Ftype - Street Type
- Leftadd1 - From Node Address (left side of street)
- Leftadd2 - To Node Address (left side of street)
- Rgtadd1 - From Node Address (right side of street)
- Rgtadd2 - To Node Address (right side of street)
- Zipcolef - Left side of street zipcode
- Zipcorgt - Right side of street zipcode
- Fromlat - From Node latitude
- Fromlong - From Node longitude
- Tolat - To Node latitude
- Tolong - To Node longitude

One of the challenges for traffic safety and GIS involves the integration of different spatial databases. Rectifying and reconciling various base maps which may have been produced

at different scales or using different map projections can be a difficult, time-consuming activity. Ideally, all the different data sources can be brought together as common layers in an integrated system. Yet often, because of the differences in the base maps, the reality may be that each system is regarded somewhat separately. One way of achieving some degree of "ground truth" and control over different spatial databases is to map data on an aerial photograph. This may involve edge matching or intersection matching or other "rubber sheeting" techniques to integrate the various databases.

Insert Figure 2.  
“TIGER Line Files - Road Network”

Filename - Tiger1.tif

By geocoding the crash files, the CODES states are able to map not just crash data, but also all of the data linked to the crash file. Since CODES link crash with EMS or emergency department, hospital, and/or insurance claims data, very valuable location information can be ascribed to crashes and the resulting injuries and costs. This type of linkage is useful at various levels of analysis. The linked data sets often contain important information that might not otherwise be available for mapping and analysis. For example, the police crash file often contains information on weather, icy roadway conditions, or other environmental data that might not necessarily be contained in the EMS or hospital databases. The EMS files, typically, contain detailed information on transport times to hospitals and emergency rooms that with GIS can help roadway planners to assess travel times and congestion levels. Additionally, these files also may have more accurate and complete information on injury outcomes, medical treatments and costs that can be analyzed by location. Potential GIS applications using linked CODES data are described below.

- ! Mapping pedestrian crashes and injury outcomes
- ! Spatial correlation of pedestrian accidents and their medical and financial outcomes with socioeconomic characteristics of neighborhoods
- ! Identification of the medical and financial outcome for hazardous roadway segments with a high incidence of run-off-the-road crashes
- ! Mapping the benefits in terms of medical and financial outcome for variations in the temporal and spatial patterns of EMS utilization along key highways
- ! Installation of traffic calming devices where they will have the most impact on reducing injuries and health care costs
- ! Mapping locations with serious injuries and high health care costs in order to prioritize installation of red-light running cameras
- ! Identification of the injury, emergency medical service, and hospital costs associated with crashes occurring at particular locations

## IMPLEMENTATION ISSUES FOR CODES/GIS

### A. GIS and Types of Map Data - Karl Kim, Hawaii CODES

It is important to know what a “GIS” entails. It can be viewed solely as a software that combines maps with a database. But a GIS can be much more encompassing. Combined with CODES, a GIS becomes a powerful system for determining spatial relationships associated with crashes.

#### ! What Is GIS?

GIS combines: people, data, software, hardware, analysis

" GIS may include different types of data:

- traffic data
- street network
- land use
- Census
- environment

" Some data in a GIS may be linked to other data such as:

- crashes linked to census
- traffic linked to EMS
- land use linked to hospital
- aerial photos linked to medical resources

#### ! What questions can a GIS answer?

" Location: *what is the worst location for pedestrian crashes?*

" Conditional: *which intersection has the most pedestrian crashes and the most traffic volume?*

" Trends: *what has changed in terms of vehicle crashes since the building of the roundabout?*

" Patterns: *are U-turns a major cause of crashes at intersections or non intersections?*

" Modeling: *what are the effects of land use change on the frequency of crashes and injury outcomes?*

## Map Properties and Types of Map Data

When establishing a GIS and before any mapping or analysis can be done, the research staff must make a decision regarding what base maps to use. When this decision is made, there needs to be some clarification and understanding about the characteristics of GIS map data. Map data come in many forms and do not always fit together, so some understanding of the principles of maps, and the problems associated with these multiple sets of data must be fully understood.

- ! What are the basic units of data used for maps?
  - " Scale: small scale (1:100,000) v. large scale (1:500)
  - " Projection: cylindrical, conical or planar
  - " Datum: NAD27 (Old Hawaiian) or WG584 (NAD83)
  - " Ellipsoid: Clarke1866 or GRS1980
  
- ! What are the basic coordinates used for location?
  - " Cartesian Coordinates
  - " Latitude/Longitude: (21 degrees 18' 01.00791" N, 157 degrees, 49' 12.65568" W)
  - " Decimal Degrees: (21.30027998 N, 157.8201821 W)
  - " State Plane: (14,804.163, 518,657.924)
  - " UTM: (2,355,838.771, 622,376,565)
  
- ! What are the map error issues, problems, and solutions?
  - " Error Issues
    - Multiple Map Databases - Within agencies there are redundancies regarding map data.
    - Multiple coordinate systems, datums & ellipsoids
    - Error propagation
  - " Problems
    - Because of the different map sources geocoded, attributes may fall off the road.
      - S A switch in datums from NAD27 to NAD83 also included the switch from the Clarke 1866 ellipsoid to the GRS1980 ellipsoid. Hence there have been excessive position shifts exceeding 250 meters in some locations.
  
- S Error propagation is defined as the means to which error accumulates and affects the end results.
  - S One source of error occurs in the geocoding process. By using past years data as the reference file for current geocoding procedures, there is the possibility that errors in geocoding may multiply. Hence, problem locations can be

over estimated.

- " Solutions
  - Standardization, by government agencies and other GIS users, of street centerline map, projection, ellipsoid, and datum.
  - Select a standard map coordinate and datum, and put all information in that form.
  - Improve accuracy of geocoding routines and develop more rigorous standards.

! What are Metadata?

- " "Data about Data": They describe the content, quality, condition and other characteristics of spatial data.
- " FGDC (Federal Geographic Data Committee) content standard for geospatial data. <http://www.fgdc.gov/metadata/metadata.html>
- " Available software: Metadata ArcView Extension  
<http://www.csc.noaa.gov/metadata/text/download.html>
- " Metamaker-USGS <http://www.umesc.usgs.gov/metamaker/download.html>

## GIS and CODES

The GIS software will not provide solutions without maps, and without data that can be associated with the maps. Because of the spatial nature of crashes, linking GIS and CODES data allows the analyst to map and determine if spatial factors are associated with crashes and injury outcomes. It is important to understand how to integrate CODES data into a GIS.

! How do GIS and CODES data work together?

- " By geocoding crash data to map databases
- " Linking CODES data to spatial databases
- " Basic spatial elements - points, lines, and polygons
- " Spatial measurement levels - interval/ratio, nominal, and ordinal
- " Spatial analysis & statistics

! What is geocoding and what are the different forms of geocoding?

- " Geocoding is the process of assigning geographical coordinates to each point, line, and area entity.
- " Point on a map - Placing a point on the map where the event has occurred.
- " Direct geocoding - Knowing the coordinate pair for the event.
- " Relative Geocoding
  - Measurement along a road network
  - Offset distance
- " Discrete Geocoding/Address matching

- Automatch/probabilistic matching
- Range matching v. corner matching

! What other types of spatial data acquisition can be used to support a GIS?

- " Global positioning systems (GPS)
  - Positional information gathering
  - Ground truth
- " Aerial photography
  - Road feature analysis and inventory
  - Importance of photo scale

! What are other spatial databases that can be linked to CODES data?

- " Census databases: *Is there a relationship between population and crashes, or income and crashes?*
- " Land use databases: *Is there a relationship between the different types of land use and the different types of crashes?*
- " Traffic analysis zone: *Is there a relationship between the number of trips made and the frequency of crash types and injury outcomes?*
- " Traffic volume: *How does traffic volume affect the number and distribution of crash types?*



## B. Development of a Traffic Safety GIS - Mary Pease, South Carolina CODES

Hardware, software and staffing configurations determine the scope and capabilities of a CODES/GIS.

- ! How do you determine what hardware will be required for a GIS Project?
  - " To assess the need for hardware for the GIS system, begin with taking into consideration GIS software requirements. While this may appear to be a "chicken or egg" type statement, the impact of software requirements must be factored into the assessment before hardware is purchased.
  - " Begin by inventorying existing hardware and software systems identifying the load and processing demands that the system can handle. If this is a new system, identify the load and processing demands using recommended software requirements, anticipated file sizes, storage capacity for archiving files and any special requirements. Special requirements might include storage for optical files or aerial files.
  - " System communications between the hardware and software may demand additional hardware requirements. Will confidentiality requirements impose requirements for the storage of multiple files, thus increasing the demand for storage space? Are there hardware requirements for security protocols that will be used on the system? What are the time requirements for producing a map?
  - " Identify existing and potential hardware and software vendors. What vendors are on contract with the state? What have been the experiences of other agencies using the hardware or software?
  - " Identifying your customers' needs (user needs) will assist in determining the hardware and software requirements for a GIS system. Do you need to have data available to large groups of people over an enterprise-wide (WAN) network? Do you need to have data/maps available to small groups where a local area network would be adequate? How will the public access information over the Internet? Do you need a secured site for access by partners or other state agencies (Extranet)? Do your customers need access to maps or do your customers need access to data to use in their GIS systems? Will the demand be to create maps from large databases on the fly or will the demand be for static maps?
- ! What factors should be considered when purchasing GIS and database software?
  - " The needs assessment for software to use with a GIS must include how the GIS software will interface with the database and/or statistical software that

will be used. Inventory existing software systems to identify the types of GIS and database/statistical software currently used by the organization. Does the database/statistical software have the capacity to handle the files anticipated for use in the system? Is the current system scalable? Can it handle the planned growth for the system? How easy is it to communicate between the GIS and the database/statistical software packages?

! What type of staff is required for a GIS System?

" The needs assessment for staff begins with categorizing the expertise of your current staff. Does anyone have GIS background? Does anyone have a statistical background? Does anyone on staff have the desire to learn or expand GIS/statistical skills? Next, determine what expertise cooperative partners may have that they are willing to share. Does anyone have GIS background? Does anyone have a statistical background? How much staff time can be shared? Will this be sufficient to meet the demands of the project? Once gaps in expertise have been identified, either on your staff and/or with partners, you must assess if additional staff is justified. What expertise is needed for GIS and/or statistics? If additional staff is required, how will the position be institutionalized after the grant funding lapses?

! How will GIS data be disseminated from the project?

" Requirements for hardware, software and staff can be affected by the type of data dissemination that a project is anticipating.

- An interactive Internet access, where people can submit files to your organization to be geocoded and returned or where they can map on the fly using a database that can be queried, requires hardware and software that can handle the number of hits anticipated and the length of time that someone will wait to run or submit a program/file. This Internet map server technology will allow users to view geographic data online and download public spatial datasets providing broad accessibility to crash information.
- Will a static Internet site meet the needs of the project? The hardware and software requirements for a static Internet site are significantly less than those for a map server that can be queried.
- How will confidentiality requirements affect the Internet site, i.e., what type of firewall will be required?
- Does the project plan to disseminate data via reports through data requests to staff or via an Internet site?

" Do not forget to plan for the future. Make data dissemination plans that can

be expanded as the project grows.

- ! How will the partners share data?
  - " Will the data reside at one location or at multiple locations? How will files be updated among partners?
  
- ! What type of database will be required for the GIS system?
  - " The type of database to use in the GIS system depends on the uniformity and data collection of the location information for the CODES project. If your project has global positioning system data available, then the base map should be in a GPS base map. If your project has address locations, then a TIGER base map may be appropriate. TIGER base maps are available for every state through the Bureau of the Census. It is important to consider the unique GIS projects that may be occurring in your state. South Carolina had a GIS project that used road updates by county tax assessors to enhance the TIGER maps. In addition, the GIS staff was able to develop a crosswalk between the address file and the GPS data.
  
- ! What mapping issues must be considered when developing a GIS System?
  - " Base maps for global positioning systems are created with one to 5 meter accuracy as compared with the TIGER maps at 50 meter accuracy. One of the problems with the GPS is that census information cannot be linked to the maps. Census data are used in analysis of many types of social and health problems. GPS to TIGER comparisons can be improved with GIS algorithms adjusting the coordinates to be correctly located on a base map, resulting in the ability to maintain topological relationships between geographic features and to correct census block assignments being established.
  
- ! What are the recommendations for selecting hardware and software and determining the staffing requirements for a GIS?
  - " Select scalable hardware
  - " Select flexible database and GIS software that can communicate with various data formats.
  - " Build a crosswalk between GIS base maps
  - " Choose data formats that can be served over the World Wide Web.
  - " Consider the advantages of enhanced location technology (GPS).
    - Accurate location information
    - Standardized data collection
  - " Hire a GIS technician and statistician
  - " Budget funds for training of technical staff
  - " Keep abreast of technology advances and trends
    - Paper analog maps

- GPS location technology
- Address geocoding

### C. Institutional Issues - Tim Kerns, Maryland CODES

Administrative factors determine who will be able to access the CODES/GIS and under what circumstances. Collaboration is necessary to balance availability, confidentiality and cost.

! Who is the 'owner' of the completed database?

" Ownership of the individual databases is maintained by the organization in charge of collecting and maintaining the data.  
The linked dataset is *maintained* by the CODES site; however, the data use is governed by the Board of Directors.

! Who controls access to the linked database?

" Currently, the CODES/GIS site will work with any organization or individual in answering highway safety questions. At the present time, the dataset itself will not be released without the specific approval of the Board. All requests for analysis will be received by the CODES/GIS site and handled in a timely fashion. All requests and the responses to them will be reported to the Board at the next scheduled meeting. Request for analyses using confidential variables must first be approved by the Board. Reports generated using CODES/GIS will be made available via the Internet.

! How is security of the data maintained?

" CDs containing CODES/GIS data are maintained in locked filing cabinets. Data stored on personal computers are maintained locally.  
No access to the raw database is currently available without prior approval from the Board of Directors.

! What safeguards are in place to preserve confidentiality?

" Confidential variables are defined by the data owners. Only non confidential variables are used in reports unless specifically permitted by the Board.

- ! What are the costs involved in the maintenance and use of the database?
  - " All currently available 'off the shelf' PCs have sufficient power to operate the software and manage the databases.
- ! What recommendations would you suggest to an interested CODES state?
  - " Follow the guidelines developed for the use of general CODES data and determine what additional variables may need to be excluded from the mapping process to maintain appropriate confidentiality.
  - " Follow the procedures outlined by your local data owners and the CODES Board for the release of site-specific information.

#### D. Linking CODES to GIS - Tom Hettinger, New Hampshire CODES

We live in a visually oriented age, where vivid graphics and computer-generated animation are the standard. The public and many policy makers are no longer satisfied to stare at rows and columns of figures... boring data. If we are to effectively communicate our findings to our audience, we must take full advantage of our current technological resources. We must add some visual appeal to our product. GIS has the ability to transform otherwise dull statistics into attractive meaningful presentations. Because we are now able to create an appealing product for public consumption, we must expect that the public will make increasing use of it. This puts us and our data in the limelight. Therefore, we must be at our best... because now, people are looking at what we have to say.

##### ! Does GIS Raise the Data Stakes?

- " If the public perceives GIS as a quick, easy information source, they are then more likely to use/abuse it.
- " Dealing with sensitive issues (DUI, drivers by age group, graduated licensing, etc.) raises the stakes. Public exposure increases and, thus, so does our need for consistency and validity in our data.

##### ! What Will We Map First?

- " Serious alcohol-related crashes.
- " Crashes involving drivers with less than one year of experience.
- " Crashes that involve drivers over the age of 75. ...*And much more.*

##### ! Is Confidentiality an Issue in a Small State?

- " Assuring anonymity is difficult with small numbers.
- " The "most restrictive" will be the rule for data access and dissemination.
- " Existing reporting thresholds are designed to prevent identification of individuals or specific incidents.
- " Confidential data may be necessary for "accepted" research projects.

##### ! How Should We Accommodate Researchers?

- " Use existing CODES data access/release policies.
- " Handle unique requests on an individual case basis.

##### ! Should GIS be Available on the Internet?

- " Step One - Static reports for viewing or downloading.
- " Step Two - Basic queries of the frequencies and cross-tabulations (above a predetermined reporting threshold).
- " Step Three - To be determined by user needs, demand and available resources.

- ! What is the Bottom Line?
  - " What we map (our *public product*) should drive the way in which we distribute it.
  - " The more likely the public is to use our CODES mapped data, the more we must focus on providing a message that is valid, consistent and unambiguous.
  
- ! Recommendations
  - " Involve all data owners in major data decisions
  - " Agree in advance on how data requests will be handled and by whom.
  - " Give data owners a useful product in return for their contribution.
  - " Focus research and analysis efforts on producing products with real world applications, which can be readily evaluated.



## E. Analytical Issues - Karl Kim, Hawaii CODES

A GIS is more than just a cartographic or map making tool. The strength of a GIS is that it provides a means of performing relational analysis between the maps, which are the spatial abstraction of the real world, and the different databases available. It is important to understand what is meant to perform spatial analysis, and what the properties and capabilities of a GIS and spatial analysis are in regards to CODES data.

- ! What is Spatial Analysis?
  - " Spatial analysis is the measurement, description, categorization, and modeling of spatial elements.
  - " Spatial analysis and statistical techniques provide better understanding of point, linear, and zonal phenomena.
  
- ! What are the basic spatial elements used in a GIS?
  - " Points - discrete spatial phenomena or event occurring in one location in space (one dimensional).
  - " Segments or lines - consisting of a beginning and an end point to describe its spatial location (one dimensional).
  - " Zones or polygons - consisting of a series of lines that begin and end at the same location (two dimensional).
  - " Need to develop a better understanding of these basic spatial elements to be able to perform analysis within a GIS environment.
  
- ! What are the spatial measurement levels used in a GIS?
  - " Nominal - level of data measurement that is non comparative, usually representing a description or name.
  - " Ordinal - ranked data that are only comparable within a given spectrum.
  - " Interval - comparative data with a relatively high degree of accuracy, but with an arbitrary starting point.
  - " Ratio - the highest level of data measurement that includes an absolute starting point and allows ratios of values to be produced.
  - " Different measurement levels provide different analysis capabilities.

- ! What are some of the spatial analysis and statistical capabilities of a GIS?
- " Proximity analysis: *What percentage of the child pedestrian crashes occur within one half mile of a school?*
  - " Spatial distribution: *Are the crashes represented by a clustered, random, or dispersed distribution? Is there a pattern in crash location?*
  - " Spatial association or correlation: *Is there a relationship between pedestrian crashes and traffic volume?*
- ! Discussion
- " Importance of map accuracy
    - Although maps are just an abstraction of the real world, it is important to use the most accurate maps available to better understand the spatial relationships between crash data and the surrounding environment.
  - " Importance of attribute information accuracy
    - Even with accurate maps, if the attribute data are not accurate, the maps become meaningless because procedures such as geocoding also are inaccurate, hence further contaminating the results of any spatial analysis.
  - " Comparability across map packages
    - With the numerous GIS software packages available, databases need to be flexible enough to be utilized in any GIS package.
  - " Need to develop better analysis methods for crashes along segments
    - There are point statistical analysis methods and zonal analysis methods but few, if any, statistical approaches to analyzing events along segments.
  - " Need for statistical and spatial statistical software
    - GIS software has not included the capability to perform statistical analysis. Thus, there is the need for the loose coupling of statistical software and GIS software.

- ! Recommendations
  - " Develop separate procedures for point, segment, and zonal analyses
  - " Use standard base maps augmented by aerial photos and GPS
  - " Use statistical packages such as SAS or Crimestat
  - " Collaborate with a geographer
  - " At the same time, recognize that a map is just one representation of reality....

STATE SPECIFIC DESCRIPTIONS

GEOGRAPHIC INFORMATION SYSTEMS

Hawaii  
Maryland  
New Hampshire  
South Carolina

## Introduction

The following state-specific descriptions use a common format to present information about the characteristics of the GIS resources, authority, funding, data elements, confidentiality and dissemination processes in Hawaii, Maryland, New Hampshire and South Carolina.

In these four CODES states, GIS was primarily developed using existing resources. Only Maryland started from scratch, involving other stakeholders as the system developed. Hawaii collaborated with the City and County of Honolulu. New Hampshire joined the E-911 efforts and South Carolina collaborated with the Department of Public Safety.

Funding sources varied. Hawaii and South Carolina obtained grants, from the University and Public Safety respectively. Maryland used CODES funds to get started. New Hampshire was funded by E-911. Currently, the GIS in all four states is a collaborative project involving stakeholders statewide.

All four states eliminated confidential identifiers and separated the data used for matching from that used for analysis. Except for New Hampshire which is still in the developmental phase, the three remaining states control direct access to the CODES/GIS data. South Carolina requires data requests to be approved by the Data Oversight Council. However, they are able to disseminate their GIS data via the Internet through a series of static and dynamic maps.

Development of a CODES/GIS has resulted in increased support for traffic safety efforts, an evaluation of pedestrian safety, creation of a state injury fact book, improved interagency communication and cooperation and improved collection of location data. Based on the experiences of Hawaii, Maryland, New Hampshire and South Carolina, a map is worth a thousand words so developing a GIS represents a significant expansion of the value of CODES.

## DESCRIPTION OF THE HAWAII GIS

! What does a GIS project really encompass?

" The Hawaii GIS project consists of hardware, software, databases and various procedures for geocoding, spatial analysis, and mapping.

! What resources did you have and what did you obtain to implement the Hawaii GIS?

HARDWARE RESOURCES	EXISTING	OBTAINED
Sun Workstation	/	
Sparc 10	/	
Ultra 60	/	
RAID Drive	/	
PCs	/	
Plotter (36")	/	
CD-ROM Writer		/

SOFTWARE RESOURCES	EXISTING	OBTAINED
Atlas*GIS	/	
MapInfo	/	
Arc/Info	/	
ArcView	/	
GeoMedia	/	
MinuTP		/
TRANS-CAD		/

DATA RESOURCES	EXISTING	OBTAINED
TIGER (1:100000 base map)	/	Needed to update the available data
DLG (1:24000 base map)	/	
Parcel Maps (1:500 base map)	/	
Crash Data*	/	
EMS Data*	/	
Hospital Data*	/	
Insurance Data*	/	
*CODES linked data geolinked to common spatial database; used Automatch to match address to map based GPS		

ANALYTICAL RESOURCES
<p>Skills:  Geocoding, map making, spatial analysis techniques, cluster analysis, nearest neighbor, spatial, auto correlation</p>
<p>Point:  Many statistical procedures</p> <p>Segment:  Fewest procedures</p> <p>Zonal representations:  Grouped data, some statistical procedures</p>
<p>Spatial statistical analysis:  Pointstat, Crimestat, SpaceStat, SAS</p>
<p>Spatial modeling:  Used to relate land use to traffic crashes</p>
<p>Accuracy, reliability, and validity of using spatial data:  Used multiple base maps, aerial photographs, and GPS to validate data</p>

! What authority controlled the process?

" The Hawaii CODES/GIS is an integral part of the CODES project located at the University of Hawaii at Manoa. It is a tool used by the social science research staff. The CODES Advisory Committee includes the GIS Coordinators for the State of Hawaii and for the City and County of Honolulu.

! What were the costs?

Developmental	We received two large grants from the University to cover hardware costs (one was for approximately \$30,000 and the second was for approximately \$60,000). We pay an additional \$4,000 per year in site licenses. The major costs were for the geocoding and spatial analysis.
Ongoing	The costs are approximately \$3,000 per month for personnel to geocode and maintain the database.

! What CODES data elements did you make accessible in your GIS?

" The crash record has been geocoded. Since the crash record is linked to EMS and hospital data, all data elements computerized in these three data files are available in the Hawaii GIS.

! What CODES data elements did you NOT include in the GIS and how were those decisions made?

" Location information on the CODES-linked EMS and hospital data files did not include the patient's address and therefore was not specific enough to be used in the geocoding. Roadway conditions, topography, land use, highway inventory, etc., were linked to the CODES data via the GIS, thus providing access to information about land use activity, facilities, and slope.



- ! How does your GIS accommodate a researcher's need for a "higher" level of access to linked data, i.e., access to variables such as dates, personal identifiers, or precise location codes?
- " After linkage, we remove the personal identifiers. We also have posted the locations to the nearest intersection rather than giving an exact location. Problems associated with the use of sensitive or confidential information are avoided by keeping the matching variables separate from the analysis variables.
- ! How do/will you disseminate your GIS data via the Internet?
- " We will distribute summaries of reports, abstracts, and maps that have been approved for publication. We have not, as yet, made shape files or GIS databases available for downloading, although some of our partnering agencies do distribute this information. We put jpeg and other image files on our site which Internet users can download.
- ! How does/will your GIS interface with other existing agencies, databases, research programs to share data and maximize effective data management/analysis?
- " We have reconfigured our CODES Advisory Committee to include participation from two of the key agencies involved with GIS: The City and County of Honolulu (Land Information System) and the GIS coordinator from the State of Hawaii. We also are partners with the regional metropolitan planning organization to share data and to conduct joint analytical exercises involving GIS, traffic safety, and transportation data.
- ! What impact has your CODES/GIS had on traffic safety so far?
- " The CODES/GIS has been an integral part of the CODES project. It provides information to those interested in GIS and to those interested in the development of GIS applications related to traffic safety.

DESCRIPTION OF THE MARYLAND GIS
---------------------------------

- ! What does a GIS project really encompass?
- " Currently, the Maryland CODES project uses GIS to calculate incidence rates and for spatial representation of crash sites.
- ! What resources did you have and what did you obtain to implement the Maryland GIS?

HARDWARE RESOURCES	EXISTING	OBTAINED
Dell Pentium III	/	

SOFTWARE RESOURCES	EXISTING	OBTAINED
MapInfo v6.0	/	
ArcView 3.2		/

DATA RESOURCES	EXISTING	OBTAINED
CODES database with street address or zip code	/	
The CODES database can be geocoded directly.	/	
MapInfo state and county maps used as base maps.	/	

ANALYTICAL RESOURCES	EXISTING	OBTAINED
Geo-spatial issues:		Currently maps are used for visual descriptions
Geocoding strategies	Manual and automatic passes	Improved address standardization

! What authority controlled the process?

" All geocoding was performed by the National Study Center at the University of Maryland.

! What were the costs?

Developmental	MapInfo software and accessories ~ \$3000
Ongoing	Software upgrade ~ \$600 annually.

! What CODES data elements did you make accessible in your GIS?

" Currently all non confidential data items are included in the geocoded

database. Demographic information (population, income, etc.) is included by census tract.

! What CODES data elements did you NOT include in the GIS and how were those decisions made?

" All personal identifiers are removed from the dataset as well as those items that are considered confidential by the individual data owners.

! How does your GIS accommodate a researcher's need for a "higher" level of access to linked data, i.e., access to variables such as dates, personal identifiers, or precise location codes?

" The current process for access to CODES data (including GIS) information is to submit a request to the agency performing the data linkage (currently University of MD/Maryland Department of Health and Mental Hygiene). Any request that involves personal or confidential information must be referred to the CODES Board of Directors and the individual data owner.

- ! How do/will you disseminate your GIS data via the Internet?
  - " Currently, there is no mechanism in place to disseminate GIS data over the Internet.
- ! How does/will your GIS interface with other existing agencies, databases, research programs to share data and maximize effective data management/analysis?
  - " All GIS analysis produced from the CODES database will be available for dissemination.
- ! What impact has your CODES/GIS had on traffic safety so far?
  - " Our CODES/GIS data have been used to support a pedestrian safety campaign in Ocean City, MD for the coming summer season. We are also using CODES/GIS data in the production of a traffic safety and injury fact book for Maryland.

## DESCRIPTION OF THE NEW HAMPSHIRE GIS

! What does a GIS project really encompass?

- " GIS provides unique perspectives from which to view data as graphical representations of frequencies and trends.
- " Focus on applying project findings toward developing real world applications.
- " Buy into the existing GIS operated by the Bureau of Emergency Communications (E-911).
- " Bring new agencies (Division of Motor Vehicles) into the CODES/GIS arena.
- " Reassess data dissemination policies and procedures in light of probable increased public appeal, availability, and usage.

! What resources did you have and what did you obtain to implement the New Hampshire GIS?

HARDWARE RESOURCES	EXISTING	OBTAINED
E-911	/	Hardware needs will grow with the system

SOFTWARE RESOURCES	EXISTING	OBTAINED
MapInfo	/	NHDOT and others use ArcInfo, so there will be some issues
ArcView 3.2	/	

DATA RESOURCES	EXISTING	OBTAINED
E-911 has street address/town roads mapping	/	CODES databases have not yet been mapped.
Linkages to CODES: We will need to develop strategies to link CODES to E-911 databases and maps.		
Base maps: E-911 has street addresses and local information as voluntarily provided by communities.		

ANALYTICAL RESOURCES
Point: Under study
Segment: Under study
Zonal representations: Under study
Spatial statistical analysis: Under study
Spatial modeling: Under study
Accuracy, reliability, and validity of using spatial data: Under study

! What authority controlled the process?

" E-911 (Enhanced 911) owns the existing GIS. However, if CODES agencies can contribute financial and personnel resources, we will have more input into the future course of the GIS.

! What were the costs?

Developmental	100% funded by E-911
Ongoing	100% funded by E-911 to this point. CODES agencies should be willing to contribute resources if we are to ask for services. This will be negotiated in the future.

! What CODES data elements did you make accessible in your GIS?

" Although still under study, we expect to include the majority of crash, EMS, and hospital data elements. We have not yet evaluated DMV data to identify desirable elements.

! What CODES data elements did you NOT include in the GIS and how were those decisions made?

" Under study at this time.

! How does your GIS accommodate a researcher's need for a "higher" level of access to linked data, i.e., access to variables such as dates, personal identifiers, or precise location codes?

" We will probably continue with our present CODES data access/release policies, which require the approval of the data owner(s) if anything more detailed than general summary reports are to be released. Researchers' requests will be handled on an individual basis. New Hampshire statutes prohibit the release of personal identifiers (name, SSN, etc.). Locations, dates, and other potential "identifiers" will be subject to the established reporting thresholds as used by the Dept. of Health & Human Services. These may be revised in the future.

! How do/will you disseminate your GIS data via the Internet?

- " Step One - Static reports for viewing/downloading
- " Step Two - Basic queries of frequencies and crosstabs
- " Step Three - Will be determined by user demands and available resources

Note: Because the Internet will make GIS/CODES data more readily available to the general public, we expect more agencies and citizens to use the system. This may require us to reexamine what we make "public" and how we present it.

! How does/will your GIS interface with other existing agencies, databases, research programs to share data and maximize effective data management/analysis?

- " Our CODES is the state's leading multi agency cooperative data sharing operation. We will continue this approach to encompass more agencies in the future and will actively support efforts to establish a statewide data warehouse. We currently do analyses and generate reports for many outside agencies and even individual citizens.

! What impact has your CODES/GIS had on traffic safety so far?

- " Although we do not yet have our links to GIS operational, the process has already further improved interagency communication and cooperation. This has probably had a positive effect on traffic safety, but there is no way to measure it at this time.



## DESCRIPTION OF THE SOUTH CAROLINA GIS

! What does a GIS project really encompass?

- " Mapping crashes
- " Defining CODES data in spatial terms
- " Enabling access to the CODES data by all users via the Internet
- " Affect policy and support grant and community development programs

! What resources did you have and what did you obtain to implement the South Carolina GIS?

HARDWARE RESOURCES	EXISTING	OBTAINED
Unix & Windows NT	/	

SOFTWARE RESOURCES	EXISTING	OBTAINED
Arc/Info	/	
ArcView	/	
Map Objects	/	

DATA RESOURCES	EXISTING	OBTAINED
SAS	/	
Crash Data*	/	*Developed linking methodology
EMS Data*	/	
Emergency Department*	/	
Hospital Data*	/	
Base Maps	TIGER	Updated from local GIS data
Geocoding strategies	Street Address street intersection	

\*\_

ANALYTICAL RESOURCES
Analytical issues What to link? What do the non matches mean?

Point: Point, zonal analysis
Geo-spatial issues (datum, projection, scale, etc.) UTM, Zone 17 Datum NAD27 1:100,000
Spatial statistical analysis: Crash frequencies by street intersection, levels (state, county, zip) of geography
Spatial modeling: Crash density maps
Accuracy, reliability, and validity of using spatial data: Accuracy: Crash location (did the crash occur where it was mapped) Reliability: Match rate of 82% Validity: The unmatched 18% - is it evenly distributed or skewed with regard to specific location variables.

! What authority controlled the process?

" South Carolina Department of Public Safety and the State Budget and Control Board, Office of Research and Statistics (ORS).

! What were the costs?

Developmental	Grant funded
Ongoing	Institutionalized into existing infrastructure with costs funded by State.

- ! What CODES data elements did you make accessible in your GIS?
- " The CODES GIS has access to all variables included in the project. They may not release information that will identify an individual; name, address, driver license number, or other identification number that may identify an individual.
- ! What CODES data elements did you NOT include in the GIS and how were those decisions made?
- " The Office of Research and Statistics entered into a Memorandum of Agreement with the Department of Public Safety to not release personal identifiers. Personal identifiers include name, address, social security number and driver license number.
- ! How does your GIS accommodate a researcher's need for a "higher" level of access to linked data, i.e., access to variables such as dates, personal identifiers, or precise location codes?
- " The Data Oversight Council is required by South Carolina law to review all requests and determine which identifying variables can be released to researchers. The decisions are based on the *Principles and Protocols for Data Release*. Under the Principles and Protocols for Data Release, only state agencies with statutory authority may receive information that identifies an individual including an individual's addresses and corresponding latitudes and longitudes. Researchers may contract with the Office of Research Statistics to perform GIS analyses using point locations with the release of data to the researcher being at a density level.
- ! How do/will you disseminate your GIS data via the Internet?
- " The Office of Research and Statistics has placed a CODES section on the Internet. GIS data are disseminated via a series of static and dynamic maps.

! How does/will your GIS interface with other existing agencies, databases, research programs to share data and maximize effective data management/ analysis?

" Additional linkages and sharing of data will be done through the Office of Research and Statistics under "the Principals and Protocol for Data Release." ORS is the central repository for all state data and health care utilization databases. ORS works with other state agencies, researchers, federal agencies and other interested parties to provide data on an as needed basis through the release of datasets or performs the work under Memorandums of Agreement. Approved data, reports and maps can be disseminated via the Internet.

! What impact has your CODES/GIS had on traffic safety so far?

" GIS has had a positive impact on the collection of the location data. The crash report form has been modified to better reflect the location of the crash. The impact has been in the technical phase of the CODES project, not yet in the programmatic phase of the CODES. The Department of Public Safety's highway grants programs are strongly urging local communities to secure maps to assist in identifying the traffic needs of their communities.