Appendix A: List of Interviewees

Environmental Monitoring Systems	Contact Person/Title/Organization
Aerometric Data Analysis and Management System	Ron Rothacker, Manager, Air Quality Data Section Michael Redgrave, Staff Air Pollution Specialist, Air Resources Board
California Emissions Inventory Development and Reporting System	Dr. Michael T. Benjamin, Manager, Emissions Inventory Systems Section, Air Resources Board
Highway Performance Monitoring System	Brian Domsic, Chief, Highway Performance Monitoring Branch, Caltrans
National Emissions Inventory Database	Phil Lorang, Emission Factor and Inventory Group Leader, US EPA Rhonda Thompson, Statistician, US EPA Lee Tooley, Environmental Scientist, US EPA
Pesticide Use Report Database	Ada Ann Scott, Data Processing Manager I, County Permit and Use Reporting Program, DPR Linda Lichtenberger, Associate Information System Analyst, County Permit & Use Reporting Program
Superfund National Priorities List Assessment Program	Terry Jeng, Environmental Scientist, Superfund NPL, US EPA
Toxics Release Inventory	Josh Woodyard, Project Officer, Toxics Release Inventory Program, US EPA
Water Quality Monitoring Database	Paul Collins, Data Processing Manager II, Drinking Water Program, DHS

Health Surveillance Systems	Contact Person/Title/Organization
Behavioral Risk Factor Survey	Holly Hoegh, Director, California Behavioral Risk Factor, Survey Research Group
California Birth Defects Monitoring Program Registry	Barbara Warmerdam, Data Operations Manager, Birth Defects Monitoring Program, DHS
California Health Interview Survey	Rick Brown, Director, UCLA Center for Health Policy Research
California Women's Health Survey	Marta Induni, Research Specialist, Survey Research Group Gary Gentry, Network Systems Engineer, Survey Research Group
Elevated Lead Visual Information System	Barbara Materna, Chief, Occupational Health Branch, DHS Susan Payne, Registry Coordinator, DHS
EUREKA (California Cancer Registry)	Steve Fuchslin, Systems Support Manager, Cancer Surveillance Section, DHS Carlos Sola-Llonch, Lead Architect, Cancer Surveillance Section, DHS
Medical Care Statistics Section System	Jim Klein, Research Specialist, Medical Care Statistics Section, DHS
Patient Discharge Database	Candace Diamond, Manager Patient Discharge Section, OSHPD Mike Kassis, Chief Information Officer, OSHPD
Response & Surveillance System for Childhood Lead Exposure	Jeff Sanchez, Chief, Health Information Systems Section, Childhood Lead Poisoning Prevention Branch, DHS Lisa Robertson, Research Associate, Childhood Lead Poisoning Prevention Branch, DHS
SENSOR Asthma Database	Jennifer Flattery, Research Scientist, Occupational Health Branch, DHS
SENSOR Pesticide Illness Database	Ximena Vergara, Research Associate, Occupational Health Branch, Agricultural Health and Safety Section, DHS John Beckman, Research Associate, Occupational Health Branch, Agricultural Health and Safety Section, DHS
Automated Vital Statistics System	Mike Quinn, Chief, Office of Information and Research, DHS

Appendix B System Summaries: Description and Key

Appendix B provides short summaries of the 20 systems surveyed based on system owner interview responses and website information. Systems are presented in the following order:

Each system summary includes:

Description and Use: A description of the system purpose, key data, data use, and primary users.

Age: The year the system was implemented and the most aged data maintained by the system.

Size: An estimated number of records maintained by the system.

Reporting Frequency: The frequency of data capture (e.g., laboratories submit reports on a daily basis).

Data Currency: Describes the most current official data available for public use.

Confidentiality: Identifies data restricted for public access.

Geographic Specificity: Denotes the most precise event location data captured by the system.

Data Collection Process: A description of system data collection activities.

QC/Reliability: Identifies data quality control checks, edits, audits and reviews.

System Support and Maintenance: Identifies current staff allocated to system maintenance and support.

Data Transfer Capability: Describes how data are shared external to the system and organization.

Planned Enhancements: Describes planned system enhancements.

IT Principals: Identifies key organization management and staff who will make decisions on data sharing and participation in the EHTP.

Data Dictionary: Indicates a "Yes" a data dictionary exists or "No" it does not.

Areas for Additional Assessment and Discussion: Topics and issues identified by Future Assessment project management team that should be addressed in Phase II.

Appendix B System Summaries: Aerometric Data Analysis and Management System (ADAM)

Description and Use: ADAM is a DB2 relational database maintained at the Teale Data Center. The system collects six million annual air quality observations with readings occurring hourly and daily. The air quality representativeness of samples varies by site and pollutant. ADAM receives a subset of monitoring data that is also collected by the Air Quality Management Information System (AQMIS). Monitoring sites and the subsequent laboratory analysis may generate dozens of data streams per site.

ADAM data is used to determine attainment and non-attainment of air quality standards, facilitate air quality control planning, and identify areas for further research.

The ARB can provide a network report or data upon request to include:

• Pollutant codes	• Type of ambient air quality monitoring	• Start hour
• Observation values	• Site identification numbers	• Assay
• Monitoring station location		

System users include Cal/EPA agency management, ARB planning and analysis staff, local air quality management districts, OEHHA, other state agencies, businesses, and consultants. The data submitters review the data and then submit data to the US EPA Air Quality System (AQS). A copy of the transactions submitted into the US EPA AQS is subsequently submitted into ADAM. ADAM provides a good response time with data indexed and partitioned.

Age	Data back to 1950's and 1960's
Size	180 million records – 6 million observations per year (recent years)
Period reported	Monitoring site reporting varies
Data Currency	Certified July 1 annually as the official record for the year, although data are collected on a regular basis.
Confidentiality	No confidential data is collected
Geographic Specificity	Monitoring site location data is precise

Data Collection Process: More than 250 air-monitoring stations in California gather data. The Air Resources Board (ARB), local Air Pollution Control Districts (APCD) or Air Quality Management Districts (AQMD), private contractors, and the National Parks Service (NPS) operate these stations. These stations house monitoring instruments that measure ambient levels of gaseous and particulate (solid and liquid aerosol) air pollutants. Monitoring data is typically received into ADAM two to four months after its occurrence. Data may change at any time after submittal if a problem is identified.

QC/Reliability: The submitters perform edit-checks on the data to ensure validity prior to or as part of submitting data to the US EPA Air Quality System (AQS) on a monthly basis. A copy of the US EPA file is also submitted into ADAM. The ARB's Monitoring and Laboratory Division audits the data collection system. Once loaded into ADAM, the data are deemed the official ambient air quality data for the State of California. Corrections to data can be done in ADAM.

System support and maintenance: A small staff performs user operations and database maintenance. The database is maintained at Teale.

Data transfer capability: Lab results and monitoring site data are submitted in a standard format. ARB distributes ADAM data on CD files generated once annually and provides extract files on special request. There is no routine data distribution or generation of customized reports. In addition, an almanac of summary information is available on the ARB website http://www.arb.ca.gov/aqd/aqd.htm.

Planned Enhancements: Future plans include enhancing web access to the air quality data.

IT Principals:	Bill Welty; Gary Knops; Ron Rothacker
Data dictionary:	None

Appendix B System Summaries: Aerometric Data Analysis and Management System (ADAM)

Areas for Additional Assessment and Discussion:

- Characterize specific factors that are collected for each data type.
- Determine whether ADAM data is used for local/neighborhood (microscale) modeling so as to produce continuous surface of criteria, stationary, and mobile source pollutants with flexibility in temporal linkage to health outcomes.
- Identify method of automating data transfer and exchange in context of newly-funded CDC pilot implementation project.

Appendix B System Summaries: California Emissions Inventory Development and Reporting System (CEIDARS)

Description and Use: The California Air Resources Board (ARB) maintains an inventory of criteria pollutant emissions, compiled by State and local air pollution control agencies. The inventory includes information on the emissions of reactive organic gases (ROG), oxides of nitrogen (NOx), oxides of sulfur (SOx), carbon monoxide (CO), and particulate matter (PM10), from three types of air pollutant sources:

- Stationary source (point source) – estimated 13,000 individual facilities with fixed point emissions data collected for 135 aggregated point source categories
- Area-wide sources – estimated 80 pollution source categories linked to group activity (e.g., consumer product use)
- Mobile sources – on road vehicles (e.g., automobiles, trucks) and off-road vehicles (e.g., trains, ships, aircraft, farm equipment)

Data, collected through source tests or a calculated estimate, are gathered on an ongoing basis and stored in the California Emissions Inventory Development and Reporting System (CEIDARS), an Oracle database. ARB and the California air pollution control and air quality management districts (35 districts) collect stationary source and area-wide source pollutant data; ARB, the California Department of Transportation, and regional transportation agencies collect mobile source pollutant data. The local districts may aggregate the area-wide pollutant data.

ARB uses the inventory data to develop the State's air pollution control program and as the basis for new regulations. A summary of the criteria pollutant inventory is published annually and includes:

• Pollutant source	• Total emissions by summary pollutant category	• Report type
• Tons emissions per day	• Total emissions by pollutant source	• Season
• Summary pollutant category	• Grand total statewide emissions	• Base year
• Pollutant		

In addition, CEIDARS data are used to develop State Implementation Plans (SIPS) and are used by Department of Toxic Substances Control (DTSC), Office of Environmental Health Hazards Assessment (OEHHA), Center for Disease Control (CDC), and the US EPA

CEIDARS integrates with toxic pollutant data (i.e., AB2588 Air Toxics “Hot Spots” Program) from the Air Toxics Emission Data System (ATEDS). The toxic pollutant inventory is updated every four years.

Age	1969 data – database in use since 1983
Size	The system includes millions of records stored in relational tables
Reporting Frequency	Annual reporting (criteria pollutants); toxic pollutants (every four years)
Data Currency	2001 data official, 2002 data due to ARB by 9/15/03
Confidentiality	Competitive information (throughput) supplied by facilities
Geographic Specificity	Emissions geo-coding is at best within 50 meters, at worst within several kilometers

Data Collection Process: Facilities develop and transmit emission estimates, electronic or hard copy, to the appropriate Air Quality Management District. Facility emission estimates reflect the average annual operating conditions. District staff may sample emissions at a facility. Electronic estimates with complete and correct data transfer seamlessly into CEIDARS.

The local districts submit data to ARB on a yearly basis. Beginning in 2003, the data are due by September 15 for the previous year. Previously, the data was due by December 31 of the year following data collection.

QC/Reliability: Age of reliable data are 10 years – older data less reliable as criteria for geocoding has changed from 1 kilometer to within 1 meter. ARB performs QA on facility estimates and transfers reliable data to US EPA. Data received by the US EPA can be several years old due to lengthy QA steps.

System Support and Maintenance: ARB has nine staff that work on data entry.

Appendix B System Summaries: California Emissions Inventory Development and Reporting System (CEIDARS)

Data Transfer Capability: 90% of the information is provided through the ARB website http://www.arb.ca.gov/emisinv/general.htm. ARB will generate custom reports upon request. Data transfer to the US EPA requires the conversion of the data to the US EPA standard format.

Planned Enhancements: ARB is developing a Community Health Air Pollution Information System (CHAPIS) to merge data from both CEIDARS and the Air Quality Management Information System (AQMIS) through a GIS/GUI public interface, with the goal to improve the quality of spatial data and precision geocoding. The goal is to record stationary points to the nearest meter. CHAPIS will be piloted in December.

IT Principals:	Contact Michael Benjamin
Data Dictionary:	Yes

Areas for Additional Assessment and Discussion:

- Elaborate on CalTrans method for estimating mobile source pollutants and on what geographic scale they are reported or maintained in CEIDARS.
- Determine whether CEIDARS data is used for local/neighborhood (microscale) modeling so as to produce continuous surface of criteria, stationary, and mobile source pollutants with flexibility in temporal linkage to health outcomes.
- Identify method of automating data transfer and exchange in context of newly-funded CDC pilot implementation project.

Appendix B System Summaries: Caltrans Highway Performance Monitoring System

Description and Use: HPMS is a national system used for the federal government (e.g., Federal Highway Administration (FHWA), Department of Transportation (DOT), and Congress) assessment of State and non-state highway systems extent, performance, condition, and operating function. At the State level, the information is used for capacity and traffic planning, road conditions, traffic analysis, and assisting with the monitoring of air quality. Caltrans manages the HPMS data collection and reporting efforts for California road and highway systems. Data are captured by individual segment, with segments defined using the agency's traffic management plan and traffic counts. The segments vary in length from 50 feet to many miles.

The HPMS was originally developed in 1978 and has had major modifications several times since its inception, most recently in 1998. Modifications in coverage and detail have been made to reflect changes in highway systems, legislation, and national priorities, reporting requirements (consolidate or streamline), and to incorporate new technology.

The information from this system is an important factor in the financial planning for roads and highways. The information is used to determine State appropriations included in the federal authorization and appropriations legislation, and the biennial Condition and Performance Reports to Congress. In addition, this information is also used to compute State allocations for inter-state maintenance, State Transportation Plan (STP), and National Highway system funding sources.

HPMS is also used for planning studies (e.g., Air Quality Monitoring). Some Metropolitan Planning Organizations (MPOs) use the data to develop the congestion management plan. Caltrans uses the data for traffic analysis, capacity and congestion management, and pavement condition analyses, assessing changes in highway system performance resulting from the implementation of highway system improvement programs. Researcher and consultants also use HPMS data. HPMS captured data includes:

• Section identification	• Section geometrics	• Section use or operation
• Highway system identification	• Traffic capacity data	• Computational factors
• Jurisdiction	• Pavement detail	• Algorithms (vehicle mile traveled)

Detailed HPMS table information is available on the web page as a PDF file.
http://www.fhwa.dot.gov/ohim/hpmsmanl/pdf/chap4.pdf

Age	Data captured in 1978; 1998 law mandated data collection
Size	46,000+ records
Reporting Frequency	Annual
Data Currency	2002 data official in October 2003
Confidentiality	Strategic Highway Network information
Geographic Specificity	Sections measured are based on textual geographic descriptions.

Data Collection Process: Caltrans Regional Offices, 476 cities, 57 counties, 16 MPOs, and other state/federal agencies (National Parks Service, U.S. Forest Service, State Parks and Recreation, Bureau of Indian Affairs) submit highway systems data (on system and off system) electronically or hard copy. HPMS excludes Bureau of Land Management data. Data are submitted on an annual basis in June.

Upon receipt, Caltrans samples and validates the submitted data and works with the submitting organization to correct the data as needed. Caltrans is required, by federal statute, to certify the public road mileage data by June 1 each year. Certified data are forwarded to the FHWA Office of Highway Policy and Information Branch. FHWA reviews the data and requests corrections as needed. In addition, Caltrans forwards official data files, tables and ad hoc reports to ARB. The US EPA receives data directly from HPMS.

QC/Reliability: Staff sample and validate record data.

System Support and Maintenance: Transportation System Information Program staff maintains the FHWA Visual Basic software and server; Caltrans IT Division manages system backup and recovery.

Data transfer capability: Data from various agencies is received and transmitted electronically. There is no direct, external access to the HPMS data.

Appendix B System Summaries: Caltrans Highway Performance Monitoring System

Planned Enhancements: Goal is to integrate HPMS with GIS tools to allow the user to select roadway information and retrieve a GIS data set.

IT Principals:	David Saiae, Chief, Office of Forecasting Travel and Analysis, 654-3330
Data Dictionary:	Yes

Areas for Additional Assessment and Discussion:

- Identify CalTrans cross-program collaboration between HPMS and GIS personnel.
- Clarify how sections are measured (e.g., textual descriptions, geographic coordinates). Focus discussion on Functional Classified Layer (FUNC) enhancement and linkage feasibility to HPMS. FUNC is a GIS coverage which links sections to actual segments that are derived from USGS Digital Line Graph (DLG) coverage (Contact Roger Ewers (916) 657-1601 or rewers@dot.ca.gov for more info).
- Identify opportunities (program and funding) for spatially and temporally enhancing traffic data for local/neighborhood scale modeling efforts in an environmental health context vis-à-vis SB702 recommendation.

Appendix B System Summaries: US EPA National Emissions Inventory Database

Description and Use: The US EPA prepares a national emissions inventory with input from numerous State and local air agencies, Indian tribes, various industry sources, and other government sources. These data are used for air dispersion modeling; regional strategy development; establishment of regulations; air toxics risk assessment; and, tracking trends in emissions over time.

The process from initial data collection to distribution of the finalized data is referred to as the National Emissions Inventory (NEI) system. The NEI system includes the collection of data via CDX, data compilation which involves merging submitted data and Quality Assuring (QA) draft versions, the NEI Oracle database which stores final versions of the detail data and summary reports, NEI ON the Internet (NEON) - a web-based distribution mechanism for US EPA Headquarters (HQ) and Regional Offices (RO) which directly queries the NEI, and other publicly available websites which house selected summary reports from the NEI Oracle database such as AirData at: http://hill.nccr.epa.gov/air/data/index.html and Envirofacts at: http://www.epa.gov/envirofw/ .

The National Emissions Inventory system captures criteria air pollutant (CAP) and hazardous air pollutant (HAP) data from as many as 60,000 point stationary sources. CAPs are identified as pollutants for which there are National Ambient Air Quality Standards (NAAQS), i.e., ozone, nitrogen oxide, carbon monoxide, particulate matter, and oxides of sulfur, due to health impacts/concerns on a regional and local scale. HAPs are generally defined as those pollutants that are known or suspected to cause serious health problems. Information on large point sources are collected every year. Smaller point source, non-point source and mobile data are collected every three years. For more information see the Consolidated Emissions Reporting Rule (CERR at: http://www.epa.gov/ttn/chief/cerr/index.html. Currently, the NEI Oracle database includes more detailed information on CAPs from 1990 to 2001 and HAPs information for 1996 and 1999.

Emissions inventory data captured by the system includes:

• Source type	• Contact information for file transmission	• Facility operating schedule data
• Specific geographic location for point sources	• Reporting period and frequency	• Annual/daily emissions for major/point facilities
• Facility emissions at point/segment/stack level	• Emission release point characteristics	• Site id and location information
• Annual and daily emissions for nonpoint sources at the county and state level	• Emission unit characteristics	

Age	Detail data – 1990 to 2001 for CAPs; 1996 and 1999 data sets for HAPs. Some aggregate information for some sources and pollutants back to 1970. National Trends information from 1970 available on spreadsheets.
Size	Approximately 100 gigabytes of annual CAP and approximately 10 gigabytes of HAP
Reporting Frequency	Annual reporting for large point sources; reporting every third year for other sources
Data Currency	CAP – year 2001; Year 2002 data to be compiled by the end of CY 2004 – an 18 month lag time
Confidentiality	None
Geographic Specificity	Captures latitude/longitude of emission release points to match to facility identifiers (e.g., street address, county center, GPS) – there may be multiple release points for a single facility. County level locations for nonpoint source emissions.

Data transfer capability: State and local agencies transfer data in NEI Input Format (NIF) through the US EPA's Central Data Exchange facility (CDX) with an electronic copy of the NEI Submittal Form. The NIF is presently the format most widely used by state and local agencies. The current version of the NIF and all user documentation and State quality assurance/quality control software is posted on the web for users to

download. The State QA software is intended for the State and local agencies to use before submitting data via CDX.

Data Collection Process: Data are transferred from the sources to the district, from the districts to the State, and from the State to the US EPA via CDX. Data compilation steps include quality control checks for format and content. Subsequent to data submittal, US EPA compiles the data to complete the national inventory, using additional data resources where and if necessary.

The data compilation part of the NEI system is where data are merged from external data sources to include the Department of Energy (DOE) power plant survey, Federal Highway Statistics Data, Trade Association data from industrial sources, and Toxics Release Inventory. For specific inventory data sectors, such as the power utility sector, the US EPA consults other federally required continuous emission monitoring (CEM) data reports in preference to data emissions from the States. In addition, changes in California's database made after data submission to the NEI may not be reflected in the NEI until much later if at all. Therefore, US EPA and California data will likely not match after a period of time.

A draft data set in NIF format of State and local files is published on an ftp site for external public review. State and local agencies, industry, and others may submit data corrections. US EPA compiles these corrections and performs final quality checks. After the detailed data has been finalized, it is then loaded into the NEI Oracle database. Data compilation processes for the 1999 NEI are detailed in the 1999 National Emissions Inventory Preparation Plan – http://www.epa.gov/ttn/chief/net/nei_plan_feb2001.pdf.

QC/Reliability: Validation and quality control efforts should occur at the State, local, and tribal agency level prior to (using the State QA tool) and after (review draft process) submission to the NEI. Quality control checks at the compilation stage include blending/merging data records to solve missing data elements; content checks to look for conditional fields, acceptable codes, numeric values within acceptable ranges; and, file and data element relationships. During the quality review process, US EPA may add records to fill in missing facilities and source categories, or to fill in missing geographic areas. Quality control activities are also dependent on the budget to perform this task and the timing for error identification. Some basic principles of the QC approach are discussed at http://www.epa.gov/ttn/chief/eiip/pm25inventory/qualdqos.html and specific checks to be performed on the 1999 NEI are outlined in the 1999 NEI Preparation Plan referenced above. Documentation discussing QA, completeness, and source information for the final 1999 Emission Inventory is at: http://www.epa.gov/ttn/chief/net/1999inventory.html

Data Distribution: After the detail data has been finalized, it is then loaded into the NEI Oracle database. Because NEON is currently only available to US EPA HQ and RO, finalized State and local files in NIF are produced and put on an ftp site similar to the draft review. Summary reports are then created, NEON is activated for that year, and selected summary reports are delivered to publicly available websites.

System support and maintenance: US EPA contracts out much of the data compilation, which includes QA, completeness, and merging of collected data

Planned Enhancements: US EPA plans to upgrade the NEI Oracle database server to increase space and query speed thereby increasing remote access to the NEI via NEON and the possibility of going public; include pre-1990 aggregate data at the most detailed level in the NEI Oracle database; merge CAP and HAP database structures; decrease the lag time for data availability; provide additional reporting functionality; and, improve data edit/update capability.

IT Principals:	Contact the Office of Environmental Information for authorization to partner.
Data dictionary:	Yes

Areas for Additional Assessment and Discussion:
None.

Appendix B System Summaries: Pesticide Use Report Database

Description and Use: Full pesticide use reporting was implemented in 1990. The PUR program collects data related to pesticides used by agricultural growers throughout the State and some non-agricultural applications. The data are used to accurately estimate dietary risk as well as exposure and potential risk to workers. Under the program, all agricultural pesticide use must be reported monthly (commercial operators within seven days) to the county agricultural commissioner who in turn reports the data to DPR. The reports include:

• Month and year of the application(s)	• Applicator name and address* (when application made by commercial operators)	• Application method (air, ground, other)
• County in which work was done	• Site ID	• US EPA/California registration number of the pesticide product applied
• Geographic location (meridian, township, range, section)	• Commodity/crop/site treated	• Pesticide product name and manufacturer
• Field location* (description)	• Acres or units planted	• Amount of product applied
• Operator ID/permit number	• Acres or units treated	• Person who prepared the report*
• Operator name and address*	• Date and time of application	

*Data not captured in state PUR database.

The PUR integrates with DPR's Product Label Database, which captures data on 40,000+ pesticide products registered for use in California. The product label database includes active and inactive product registrations. Throughout the year, DPR adds newly registered products, inactivates products, and processes label amendments. All use report data are run against the product label database as a check for accuracy and to convert the pounds of product applied to pounds of active ingredient(s).

DPR staff, the public, registrants, county agricultural commissioners, poison control centers, the state Legislature, and other governmental agencies use PUR data available on DPR's web site http://www.cdpr.ca.gov/docs/pur/purmain.htm.

Age	1990*
Size	An average of 2.4 million records per year (A "record" means the application of one pesticide product and all associated data recorded on the reporting form. A pesticide "application" can be a single product or a combination of multiple products, such as a tank mix.)
Reporting Frequency	Varies – monthly, quarterly from counties
Data Completeness	1990 – 2002
Data Currency	FY 2001 official data (FY 2002 available October 2003)
Confidentiality	None
Geographic Specificity	Section for geo-coding is one square mile

*Prior to 1990, a limited pesticide use reporting program was in place; only those pesticide products applied by commercial applicators and restricted use products applied by growers were required to be reported to DPR. Annual database size was approximately 850,000 records maximum.

Data Collection Process: County Agricultural Commissioners enter (data entry or electronic receipt via modem or web) records into the County Use Report Database. The county validates entered data and periodically downloads data to an electronic file to transfer to DPR's Pest Management and Licensing Branch via email. Beginning with 2002, all data was entered by the counties and submitted electronically to DPR. Due to resource constraints, DPR currently has a 7-month backlog of processing and review of counties data. Under ideal conditions, this delay would be reduced to four months (one month to submit data, 3 months for county email quarterly).

QC/Reliability: The PUR conducts up to 50 different validity checks against the data, to include verification of product registration number against legal use of reported commodity, and an outlier program to identify unlikely use rates and flag these suspect records. DPR forwards data entry error lists to the counties for their

review, research and correction. Error rates are below 2%. With the exception of duplicate records, all records are transferred to the PUR table. Fields with errors are replaced with null values or estimates. All records with errors are linked to an errors table that contains information about the errors.

System support and maintenance: Minimal support (2 PYs) for ongoing system support and maintenance including PUR data collection, processing and coordination with counties. No support for external analyses and data interpretation.

Data transfer capability: The full database and supporting documentation can be purchased on CD-ROM for all years beginning with the 1990 data. The CD-ROM is available with the data in ASCII text comma delimited format. An additional arc interchange (.e00 extension) CD-ROM with spatial representation of statewide Public Land Survey System sections is available to link pesticide use data with GIS software programs. This CD-ROM makes spatial display and analysis possible. In addition, requests can be made for specific data, such as all reports for one county, one commodity, and specific section-township-range, in any combination of data fields. These special queries can be made at DPR's California Pesticide Information Portal (CalPIP) website at http://www.calpip.cdpr.ca.gov/cfdocs/calpip/prod/main.cfm.

Planned Enhancements: Approximately 35 counties have or are in the process of digitizing (geo-coded) field boundaries (sites) that would allow more precise location reporting. Limited county resources and funding are deterrents for many counties to move forward with GIS development. DPR is completing a feasibility study report that would re-engineer the statewide permitting and use reporting system (SPURS) by developing a secured and centralized web-based application. Two primary goals of this project are to fully integrate GIS capabilities into the daily permitting program and to improve PUR data quality, integrity and reliability by pushing all data validations to the counties and users (applicators and growers). Funding for this project has not yet been determined.

IT Principals:	Ada Ann Scott, Manager Larry Wilhoit, System Coordinator Linda Lichtenberger, Data Coordinator Technology Investment Review Committee (TIRC) Technology Initiatives Leadership Team (TILT)
Data dictionary:	Yes

Areas for Additional Assessment and Discussion:

- Identify potential sources of collaboration (program and funding) to enhance PUR for additional data to support EHTN, i.e., facilitating the digitization of fields for all CA counties.
- Identify method of automating data transfer and exchange in context of the newly funded CDC pilot implementation project.

Appendix B System Summaries: Superfund National Priorities List Assessment Program

Description and Use: The US EPA uses a four-part process to estimate the chance that contact with chemicals from a site will harm people now or in the future. The activities include data collection and evaluation, exposure assessment, toxicity assessment, and risk characterization. Each of these activities provides information to determine a Hazard Ranking System (HRS) score. Those sites deemed as uncontrollable waste sites and have HRS scores above 28.5 are proposed for placement on the National Priorities List (NPL). Factors used to determine risk include:

- The likelihood that a site has released or has the potential to release hazardous substances into the environment
- The characteristics of the waste (e.g., toxicity and waste quantity)
- The people or sensitive environments (targets) affected by the release

The US EPA updates the Superfund National Priorities List Assessment Program (SNAP), using an Oracle database, with proposed site information and tracks the site selection status.

The US EPA publishes notices in the Federal Register, listing which sites are being proposed to the NPL, providing a 60-day public comment period. Sites are most often referred by States seeking Superfund support for remedial and removal programs. Placement on the NPL initiates the site cleanup process (e.g., investigation, development of a cleanup plan, physical construction for cleanup, cleanup), with cleanup performed by US EPA contractors or through US EPA oversight of independent contractors.

Congress, federal agencies, States, and the public use SNAP data. SNAP provides annual reports to Congress, and reports and analyses upon request. SNAP data includes:

• Contaminated media	• Cleanup activity specifics	• NPL status
• Contaminants of concern	• Site location information	• CERCLIS ID
• Site identifiers	• Spill identifiers	• US EPA Region
• Contaminant date	• Federal Register dates and citations	• Construction and cleanup completion dates

For a complete listing of SNAP data elements see http://www.epa.gov/superfund/sites/ded/index.htm.

Age	Estimate development in 1990; data from the 1980's added to SNAP
Size	1200+ sites or records; California has 96 sites
Reporting Frequency	Quarterly publication of proposed sites in the Federal Register
Data Currency	Site information is official for proposed locations that have been noticed and for which all public comments have been addressed
Confidentiality	All data are publicly available
Geographic Specificity	Captures latitude/longitude; street address

Data Collection Process: Sites are identified and proposed for selection on an ongoing basis, presented in "batches" within quarterly Federal Register notices. Sites may remain proposed to the NPL for 3 months to years; often dependent upon the amount of time to respond to all comments received upon notice. US EPA staff enters site information as it is identified.

QC/Reliability: A contractor performs quality control on SNAP records. This includes use of data characteristics which summarize information in the Hazard Ranking System and the identification of information inconsistent with record detail.

System Support and Maintenance: A contractor supports and maintains SNAP.

Data Transfer Capability: There is no direct access to the data but US EPA will provide the data to anyone upon informal request.

Appendix B System Summaries: Superfund National Priorities List Assessment Program

Planned Enhancements: No plans to expand the system. Proponents for the Comprehensive Environmental Response, Compensation and Liability Information System (CERCLIS) have suggested the use of CERCLIS only, in lieu of SNAP.

IT Principals:	Terry Jeng, Environmental Scientist, US EPA
Data Dictionary:	Yes

Areas for Additional Assessment and Discussion:
None

Appendix B System Summaries: US EPA Toxics Release Inventory Program

Description and Use: The Toxics Release Inventory (TRI) is a US EPA Oracle database that contains information on toxic chemical releases to the air, water, land and injected underground. The database also tracks the quantities of toxic chemicals that are transported to waste management facilities for further handling. Stakeholders do not access the database. Instead, stored information is sent to TRI Explorer (www.epa.gov/triexplorer) and Envirofacts (www.epa.gov/enviro) for public viewing.

The inventory, established under the Emergency Planning and Community Right-to-Know Act of 1986 (EPCRA), requires annual reporting of catastrophic or other one time releases by media type for an estimated 650 chemicals and chemical categories from industries including manufacturing, metal and coal mining, electric utilities that combust coal and/or oil, hazardous waste treatment and disposal facilities, chemical wholesale distributors, petroleum bulk stations and terminals, solvent recovery services, and federal facilities. These facilities must report is they employ 10 or more full-time employees and/or manufactures and processes more than 25,000 pounds of chemicals or uses more than 10,000 pounds of any listed chemical during the calendar year.

The Pollution Prevention Act of 1990 expands the reporting requirement to include waste management and source reduction activities (e.g., chemicals treated on site, recycled, combusted for energy recovery).

Since its inception in 1987, the US EPA has doubled the number of chemicals reported to the current 650 chemicals, and added new industry sectors. In addition, the US EPA recently reduced reporting thresholds for specific persistent, bioaccumulative, and toxic chemicals. TRI data includes:

• Chemicals stored on site	• Facility identifying information	• Total State releases
• On site release (pounds)	• Contact person information	• Offsite transfer to disposal
• Offsite release (pounds)	• Environmental permits	• Facility activities involving chemicals
• Environmental media	• Chemicals burned, treated, or recycled at facility	• Source reduction activities

The data are used by a variety of stakeholders including academia, research organizations, industrial associations and environmental justice advocates. Government organizations use TRI data to identify and target industries of concern. The inventory has resulted in a significant reduction of chemical releases from the reporting industries. Data are available to the public to understand hazards within their local environment. For additional background information, see http://www.epa.gov/tri/tridata/tri01/press/FactorsToConPDF.pdf. The TRI website also provides access to specific data and trend information on individual facilities, counties, states, or the nation as a whole. In addition, the data has been used to provide an analysis of toxic release data by industry, specific media (e.g., air, water, or land), and reported chemical.

Age	1987
Size	1+ million records
Reporting Frequency	Annual
Data Currency	Facilities submit calendar year reports by July of the following year; official data released up to 18 months following year end reporting period
Confidentiality	Trade secret information is not released
Geographic Specificity	Facility location information to include latitude and longitude

Data Collection Process: Reports for each calendar year are due July 1 of the following year. TRI distributes a TRI Made Easy (TRIME) software program with instructions to reporting facilities. 92% of reporting facilities are using this software program. TRIME allows the users to submit their report (Form R) on paper, floppy disk, or electronically via an Internet connection to both TRI and the appropriate State agencies. Once the form is submitted to TRI, the facilities may also submit to the respective State Agency (California only reports the Form R information to TRI).

Appendix B System Summaries: US EPA Toxics Release Inventory Program

Agencies may update release records retroactively to the 1987 reporting period. For example, if a chemical is pulled from the list of reportable chemicals, a facility may revise previous records to exclude reported release of that chemical. The system tracks record revisions.

QC/Reliability: Quality control activities include TRIME software validations and TRI expert review of data to identify and investigate under or over reporting. If potential errors are identified, the facilities are notified to investigate and correct any reporting errors. After completion of data entry and data quality activities, the TRI data are available in printed reports, in a computer database, and through a variety of informational products. States also make available to the public forms filed by individual facilities in their jurisdiction.

System Support and Maintenance: TRI contracts system support and maintenance and has 30+ contractors filling roles ranging from mail clerk to database administrators.

Data Transfer Capability: US EPA compiles the TRI data each year and makes it available through several data access tools, including the TRI Explorer and Envirofacts. There are other organizations, which also make the data available to the public through their own data access tools, including Unison Institute, which has developed a tool called "RTKNet" and Environmental Defense, which has developed a tool called "Scorecard."

Planned Enhancements: There are no known enhancements planned for the TRI system

IT Principals:	Tim Antisdel, Alternate Assignment Manager; Josh Woodyard, Project Officer
Data Dictionary:	Yes

Areas for Additional Assessment and Discussion:

- Identify specific areas of overlap with State emission inventories and determine a method for extracting uncaptured elements or elements having existing element-specific attributes beyond that which is captured in State systems.
- Identify possible methods of ongoing automated data exchange in the future.
- Identify additional programs that have used TRI data for environmental health purposes.
- Determine whether geographic coordinates are accurate enough to be used in local/neighborhood scale modeling of exposure.
- Clarify whether this data is collected by environmental media type.

Appendix B System Summaries: Water Quality Monitoring Database

Description and Use: Implemented in 1986, the Water Quality Monitoring Database collects information on the quality of drinking water from over 16,000 active wells throughout the State. These wells provide water to more than 6,000 public water systems in California. The Water Quality Monitoring Database maintains the results from water testing laboratories on over 130 chemicals monitored through periodic testing of the wells. Data includes sample of measurements, before and after water treatment.

The data are used by a variety of stakeholders, both internal and external to DHS. The DHS Division of Drinking Water and Environmental Management access the data to enforce federal and California Safe Drinking Water Acts (SDWAs) and to oversee public water systems to assure the delivery of safe drinking water to all Californians. Violators are tracked in a separate system. The data are also used to support the activities of other DHS divisions (e.g., research impact of water quality on public health outcomes), the Cal/EPA (e.g., State Water Resources Control Board, environmental studies), and respond to ad hoc requests from environmental groups, the press and other organizations with an interest in water quality.

The system captures information for monitoring and reporting purposes that includes:

• Water system identification information	• Text description of water system location and status	• Location of sampling station (well, surface water source, before/after treatment, distribution system sample)
• Water treatment information (process and objective)	• Sampling date and time	• Constituent (chemical) analyzed
• Laboratory completing the analysis	• Date of analysis	• Results of analysis

The location of the well is captured, and geo coded to within 5 meters. This information is confidential and disclosed only to government organizations with a non-disclosure agreement.

The Water Quality Monitoring Database does not collect information on the distribution points from the wells, or the delivery mechanisms of the water. Water systems may rotate usage of specific wells or blend them into a distribution source.

Age	1986; includes data back to 1984
Size	15 million records
Reporting Frequency	1000's of records reported daily
Data Currency	Data are updated daily with ongoing QC checks; data are official within two months of test
Confidentiality	Source (well) locations are confidential; share information using agency disclosure forms
Geographic Specificity	Required by law to identify latitude and longitude within 5 meters of source; estimate current data are within 1 meter

Data Collection Process: Public Water systems contract with laboratories to test source or well water quality. There are an estimated 250 laboratories in California performing and reporting on source water quality analyses. The laboratories perform an analysis and submit the results to the Water Quality Monitoring Database. The sample must be submitted by the 10th day of the following month it was analyzed. The laboratories complete the analysis and submit using a State developed system (Write-On), or similar system. The files are submitted to the State via email following a fixed format Electronic Submission Specification (ESS) where they are input into the Water Quality Monitoring Database. Automated edits are performed to check the data for validity.

QC/Reliability: Three full-time staff receive, enter and perform quality checks on reported data. The system also performs edits to ensure data accuracy. The system has a high degree of accuracy; an estimated 99.9% of records are complete and available in the system.

System Support and Maintenance: The Drinking Water Program has an IT staff that supports the database maintenance.

Appendix B System Summaries: Water Quality Monitoring Database

Data Transfer Capability: The Water Quality Monitoring Database data are available on a CD and updated on demand. Summary formats are available on the Internet for specific chemicals of current interest. Ad hoc requests may be submitted to develop reports or create files for download per the Public Records Act.

Planned Enhancements: The Drinking Water Program plans to incorporate an XML reporting format and to transition to the use of small servers vs. the current mainframe platform.

IT Principals:	Paul Collins, Data Processing Manager II, DHS Division of Drinking Water and Environmental Management
Data Dictionary:	Yes

Areas for Additional Assessment and Discussion:

- Discuss possibilities to improve treatment and sampling station network information such that water quality can be estimated/modeled for service areas downstream from sampling stations. Identify areas of collaboration to further this purpose.
- Identify possible methods of ongoing automated data exchange in the future.

Appendix B System Summaries: Behavioral Risk Factor Survey

Description and Use: The California Behavioral Risk Factor Survey (BRFS), partially funded by CDC, has been ongoing since 1984. The emphasis of this survey is on health-related behaviors in the California adult population with a specific focus on risk behaviors related to health, chronic disease, and injury. The survey is part of a surveillance effort conducted by the California Department of Health Services and the Public Health Institute's Survey Research Group (SRG) in cooperation with CDC. The annual sample size for this survey is approximately 4,000 interviews. Data collected includes:

• Health status	• Routine checkup	• Smoking
• Health insurance	• Diabetes	• Pregnancy
• Women's health	• HIV/AIDs	• Demographics
• Hypertension	• Injuries	• Alcohol use
• Vaccinations	• Colorectal screening	• Cholesterol
• Physical activities	• Fruit and vegetable consumption	• Weight control

Data are provided to State and local health departments, academic researchers, nonprofit organizations, managed care organizations, students, and occasionally the general public. Data results are used to assess intervention programs and to analyze trends in health behaviors. The data are also used to publish scientific articles, and educate the public, the professional health community and policymakers about the prevalence of modifiable behavioral risk factors and of preventive health screening practices.

For more information see http://www.cdc.gov/brfss/ and http://www.surveyresearchgroup.com/clients.asp?ID=9.

Age	Survey has been ongoing since 1984
Size	Approximately 4,000 respondents per year
Reporting Frequency	Annually
Data Currency	Data are released in March of each year for the previous calendar year. Unweighted quarterly frequencies are provided throughout the year
Confidentiality	Released data sets exclude population identifiers, zip code and county for those counties with less than 100,000 population
Geographic Specificity	County

Data Collection Process: Data are collected using interviewing software, with interview results exported into a SAS dataset for analysis and quality control checks.

QC/Reliability: A Public Health Institute Research Scientist reviews interview results. Interviewer Supervisors review data collection and provide feedback and training to the interviewers. Supervisors call 5-10% of those surveyed to verify responses. In addition, interviews are monitored for quality a minimum of 8 times per month.

System support and maintenance: One staff that also performs quality control maintains the survey data.

Data transfer capability: Data are available upon request. Annual data sets are produced in SAS format.

Planned Enhancements: There are no plans to expand the survey.

IT Principals:	Bonnie Davis, Unit Chief, Survey Research Group
Data dictionary:	Yes

Appendix B System Summaries: California Birth Defects Monitoring Program Registry

Description and Use:
The State is mandated to collect data for all of California on birth defects, monitoring patients from birth to age one. Due to budget constraints, data are currently collected from San Diego County, Orange County, two Bay Area counties, Los Angeles County, and Central Valley counties. The Department of Health Services and the March of Dimes jointly administer the Birth Defects Monitoring Program (CBDMP).

The CBDMP estimates case counts statewide and for each county by linking data from the CBDMP's Registry to Vital Statistics records. These data are used to document birth defects' public health impact, identify risk and protective factors, and to respond to the public's concerns. For more information, see http://www.cbdmp.org/spd_history.htm

Data captured within the CBDMP Registry includes:

• Patient identification information	• Lab test results	• Family history data
• Diagnoses and medical condition	• Environmental exposures	• Medical provider information

Scientists, health care professionals and the public use the Birth Registry data.

Age	Data available to 1982; Online abstraction system developed in 1995
Size	Estimated 104,000 records; data back to 1982
Period reported	Six month periods
Data Currency	7 month lag time for data to be official after patient discharge and diagnosis eligibility period is first 12 months of life, final dataset available 4-6 months after the 12 month surveillance period ends
Confidentiality	Patient data are restricted; requires approval for release from the Committee for the Protection of Human Subjects
Geographic Specificity	Patient/parent address data

Data Collection Process: To collect data, California Birth Defects Monitoring Program (CBDMP) staff visit hospitals with maternity and pediatric services and. Staff review maternity and pediatric floor logbook admitting diagnoses and hospital diagnostic discharge indexes to identify potential reportable cases (e.g., structural birth defects, chromosome abnormalities, birth defect patterns such as fetal alcohol syndrome). Staff review the medical records to include eligible diagnoses for each reportable case and use laptops to enter case information at the hospital site into a FoxPro program. This information is later uploaded into a central FoxPro database. Data are also collected from genetic offices and genetic laboratories.

The central FoxPro database performs record validation checks and all records undergo a Supervisory review. Valid records are matched to the Automated Vital Statistics System (AVSS) master file and the Fetal Deaths System master file; specific data elements in the Birth Defects Registry are replaced by data from these systems.

QC/Reliability: The Birth Defects Registry data set has varying eligibility criteria and the geographic area for data collection has changed over time.

Uploaded data are validated by the system to identify duplicate cases and other errors. System quality control is followed by a Supervisory review of all records and Supervisory audits of select cases to include re-abstracting records. CBDMP staff perform a second layer of quality control checks following the Supervisory review.

System support and maintenance: The system is supported and maintained by CBDMP programming staff.

Data transfer capability: Abstractors collect information and upload to the primary CBDMP system. There is no direct, external access to the database.

Appendix B System Summaries: California Birth Defects Monitoring Program Registry

Planned Enhancements: There are no known enhancements planned for the system.

IT Principals:	Dr. John Harris, Program Chief; Dr. Gary Shaw, Research Director; Mary Jo Campodonica, Data Director
Data dictionary:	Yes

Areas for Additional Assessment and Discussion:

- Identify feasibility of, and resources for collecting family residential history.
- Identify possible methods of ongoing automated data exchange in the future.
- Determine whether GIS web/application services provided by EHTN or CDC would be desired and could be incorporated into existing reporting application architecture.

Appendix B System Summaries: California Health Interview Survey

Description and Use: The California Health Interview Survey (CHIS) is a collaborative project between the UCLA Center for Health Policy Research, the California Department of Health Services, and the Public Health Institute to conduct a population-based survey. The biannual survey was first conducted in 2001; the most current official data available to the public is for that survey year. This survey completed interviews with 55,000 households selected at random from every county in the state. The sample is stratified into 41 geographic strata with the 33 most populous counties forming independent strata (the cities of Pasadena, Long Beach and Berkeley, which have their own health departments, were also separately sampled in CHIS 2001). The remaining 25 counties were aggregated into eight separate sample groups. There was also a supplemental sample of 2,500 households representing key ethnic groups. The 2001 survey resulted in approximately 75,000 respondents including 57,000 adults, 6,000 adolescents, and 12,000 children. The 2003 survey is currently in the field, with data collection expected to end in January 2004 and data available in the latter half of that year.

The biannual survey collects information on a wide variety of topics including health status, information on chronic illnesses such as asthma, health related behaviors, health insurance coverage, and access to, and utilization of, health care services. One adult (age 18 and older) is randomly selected from each household for the adult interview. If the adult respondent is the parent or guardian of either a child (age 0 through 11) or teen (age 12 through 27) living in the household, additional interviews are conducted. The teen interview is conducted directly with the selected teen following parental permission. For children, the adult most knowledgeable about the child's health is interviewed about the health of the selected child. The survey has the ability to be modified to collect specific health related information to support specific health related activities. In addition, at the end of the survey, respondents may be recruited to participate in specific follow-up panels based on respondent characteristics identified during the survey.

Information from the survey provides valuable data to State and local government programs, providers, foundations and other organizations that support general, or targeted health services. The CHIS provides public use files for research purposes, and supports the "*Ask*CHIS" application on the website that responds to ad hoc data requests. More information on CHIS, and the *Ask*CHIS application are available on the web site http://www.chis.ucla.edu.

Information collected through the CHIS survey includes:

• Demographic information of respondent (age, sex, race/ethnicity)	• Geographic descriptors (urban, rural, city, county, zip code, and cross-streets for Los Angeles and San Diego Counties. Street address or cross-street will be collected for all CHIS 2003 households)	• Health status
• Health conditions	• Health behavior indicators	• Functional limitations (disabilities)
• Health care coverage	• Income strata	• Women's health

The data are stored on a server located at the UCLA Center for Health Policy Research. Person identifiable data (e.g., name, telephone number, address) are not stored in the database. Public use files also exclude city, county and zip code information.

Age	2001
Size	75,000 respondent records from 2001 survey
Reporting Frequency	2001 survey
Data Currency	2001 survey results
Confidentiality	Patient identifiers are not collected
Geographic Specificity	County, city, and zip code

Data Collection Process: The information is collected through a random-digit dial (RDD) computer assisted telephone interview (CATI) survey.

Appendix B System Summaries: California Health Interview Survey

QC/Reliability: The data collected is person reported information. The CHIS uses a contractor, Westat, Inc., to conduct the survey. Westat incorporates internal training and quality control measures within their survey management methodology.

System support and maintenance: The data are maintained at the CHIS site at UCLA.

Data transfer capability: Currently, there is a public use file that has indirect identifiers (city, county, and zip code) removed. The public use file is available through the CHIS web site or CD. CHIS staff can also respond to ad hoc requests, based on appropriate authorization and documentation of the request. The California Dept. of Health Services, and other CHIS funders (e.g., National Cancer Institute), receive files that include most indirect identifiers under data sharing agreement.

Planned Enhancements: In addition to updating questions and topics in the survey, information will be collected related to geographic location of the respondent (address, nearest cross streets). Formal appeals to the Director are required if any data linkage project is undertaken with CHIS.

IT Principals:	Rick Brown, PhD, Principal Investigator CHIS; David Grant, PhD, Survey Operations Manager
Data dictionary:	Yes

Areas for Additional Assessment and Discussion:

- Identify feasibility of, and resources for collecting family residential history.
- Identify possible methods of ongoing automated data exchange in the future.
- Determine whether GIS web/application services provided by EHTN or CDC would be desired and could be incorporated into existing reporting application architecture.

Appendix B System Summaries: California Women's Health Survey

Description and Use: The California Women's Health Survey (CWHS) collects information on a variety of health behaviors and attitudes from a sample of randomly selected adult women. The annual population based telephone survey is conducted by the Survey Research Group of the Public Health Institute and is supported by a consortium of DHS programs, California Department of Mental Health, California Department of Drug and Alcohol Programs, Department of Social Services, and CA Medical Review, Inc. The survey has been ongoing since 1997 with no planned end date. Data are collected through a computer-assisted telephone interview in which approximately 4,000 women, age 18+ are randomly selected to participate. The survey consists of approximately 200 questions related to a variety of health topics. Response rates vary from 67% to 81%.

The data collected from the survey includes:

• Reproductive health	• Access to health care	• Insurance status
• Maternal and child health	• Mental health	• Nutrition
• Domestic violence	• Disability and chronic pain	• HIV testing
• Aging women	• Caregiving	• Sexually transmitted diseases

Data are provided to all consortium participants. If requested, researchers, academia, and the public may also receive the dataset. In addition, the information is used to assess intervention programs and analyze trends in women's health. For more information see http://www.dhs.ca.gov/director/owh/html/whs.htm or www.surveyresearchgroup.org.

Age	1997
Size	24,218 records
Period reported	Daily
Data Currency	Data are for CY 2002; requires one quarter to review and make official an annual data set; data frequencies are provided throughout the year as unofficial data
Confidentiality	Released data sets exclude population and is aggregated by County
Geographic Specificity	Zip Code

Data Collection Process: Data are collected using interviewing software, with interview results exported into SAS for analysis and quality control checks. Following the quality control review, a flat file is imported into the server where data are stored in an Xbase format on the server.

QC/Reliability: A Public Health Institute Research Scientist reviews interview results. Interviewer Supervisors review data collection and provide feedback and training to the interviewers. Supervisors call 5-10% of those surveyed to verify responses. In addition, interviews are monitored for quality a minimum of 8 times per month.

System support and maintenance: The survey data are maintained by one staff who also performs quality control.

Data transfer capability: Data are available upon request. Annual data sets are produced in SAS format.

Planned Enhancements: There are no plans to expand the survey.

IT Principals:	Bonnie Davis, Unit Chief, Survey Research Group
Data dictionary:	Yes

Areas for Additional Assessment and Discussion:
None

Appendix B System Summaries: Elevated Lead Visual Information System (ELVIS)

Description and Use: Since 1987, laboratories performing blood lead level (BLL) analysis on samples drawn in California have been required to report BLL results to CDHS. The Occupational Lead Poisoning Prevention Program (OLPPP) shares a surveillance system for lead poisoning with the Childhood Lead Poisoning Prevention Branch (CLPPB; see also RASSCLE system description).

From 1987 through 2002, only elevated (>= 25 ug/dl) BLL results were required to be reported; since 01/01/03 all BLLs are required to be reported. The Elevated Lead Visual Information System (ELVIS) tracks BLLs and related information for adults (ages 16+). ELVIS receives approximately 25,000 reports annually, for some 20,000 persons. Currently, approximately 70-80 workers receive case management from OLPPP annually, including approximately 15-20 who must be removed from work with lead under Cal/OSHA standards. An additional 150 workers are referred to local health departments, who send additional material to encourage family testing. Case management by OLPPP includes ensuring removal from lead exposure, appropriate medical care, and prevention of other poisonings resulting from the same exposure. OLPPP works cooperatively with employers to reduce workplace lead exposures, but if necessary, will refer an employer to Cal/OSHA for potential enforcement action.

Along with identification of cases requiring case management, ELVIS tracking data are used to ensure that required testing is being performed, to identify industries that need educational interventions, and to target OLPPP's prevention work.

BLL results must be submitted to CDHS within 3 working days if the BLL exceeds 10 ug/dl; results < 10 ug/dl must be submitted within 30 days. If a laboratory identifies a highly elevated (>= 45 ug/dl) blood lead level, they are asked to immediately fax the results to the State.

Information reported from laboratories and stored in ELVIS includes:

• Patient identification information	• Employer information	• Test result and laboratory information
• Physician identification information	• Facility of physician ordering test	• Date of blood draw
• Date lab analyzed	• Date lab receives sample	

The State reports quarterly to the CDC with an electronic export of data in CDC report format. The CDC receives reports and aggregates data from the 35 states with adult lead monitoring programs.

Age	1987 - present
Size	143,000 lab reports; 25,000 records added annually
Reporting Frequency	Daily
Data Currency	Data received and entered daily; official data within two months; continual updates to prior data; continual QC to ensure accuracy
Confidentiality	Restricted release of personal and medical information
Geographic Specificity	For all samples collected in CA, report information includes patient address, employer address, and physician address

Data Collection Process: Laboratories submit test results in either paper format (40%) or electronically (60%). DHS staff review, edit, and audit report findings to identify cases, which require follow-up or referral.

QC/Reliability: BLL reports include information available at the time of the test and may include missing fields. CDHS is working with medical providers, draw stations, and referring laboratories to improve transmission of complete information passing between them, and thus to improve the completeness of BLL reports received by CDHS.

QC is performed on both paper and electronic reports, and ensures that duplicate results are not entered, that information is accurate and complete, and that the correct person, physician, employer, and laboratory is associated with each result received. Automated validation checks occur with data entry, and periodic review of entered data also ensures QC.

Appendix B System Summaries: Elevated Lead Visual Information System (ELVIS)

System Support and Maintenance: Currently, one staff enters report data, one staff prepares electronic files for entry into the system, one staff performs a quality control check of this entry, and 1.5 full-time staff maintain the system.

Data Transfer Capability: ELVIS receives results electronically from text files that are output by a system set up by CLPPB. In the future the data transfer system between laboratories and CDHS will be web-enabled (see below). There is currently no direct remote access to the FoxPro database.

Planned Enhancements: ELVIS will be upgraded to FoxPro v 8. All results will be received electronically by 01/01/05. A web-based electronic data entry system will be enabled. Additional case management functions will be automated with the upgrade.

IT Principals:	Barbara Materna, Chief, Occupational Health Branch; Susan Payne, Registry Coordinator
Data Dictionary:	Yes

Areas for Additional Assessment and Discussion:
None

Appendix B System Summaries: EUREKA (California Cancer Registry)

Description and Use: The California Cancer Registry (CCR) is the organization authorized to receive and collect cancer data and maintain this data within a statewide population-based cancer surveillance system, referred to as EUREKA. The CCR collects information about all cancers diagnosed in California (with a few exceptions). Information is received by regional cancer registries from hospitals and pathology laboratories related to admissions for cancer. The data from regional cancer registries are combined within the EUREKA system, to support the CCR information requirements and to provide a comprehensive statewide database.

The availability of data from EUREKA allows health researchers to analyze geographic, ethnic, occupational and other differences to identify risk factors. The data also helps determine where early detection, educational or other programs should be directed. EUREKA is an important component of the CCR's effort to support additional research projects.

Information collected in EUREKA includes:

• Patient identifier information	• Patient demographic information	• Extent of disease at diagnosis
• Hospital admission data	• Attending physician information	• Tumor information
• Cancer type	• Vital status of patient	• Treatment

EUREKA data are stored in an SQL Server 2000 database. The system produces standard reports, and CCR staff have the ability to respond to ad hoc queries. Two data marts are created from the EUREKA database. One of the data marts supports standard reporting and the other provides extracts for research purposes. The CCR has developed a comprehensive process to review, approve, and track the release of confidential data from the system.

Age	Current system (EUREKA) developed in 2002
Size	Data on an estimated 4 million patients
Reporting Frequency	Converted data from 1988
Data Currency	Hospitals submit records within 6 months of admission, and the review and update process may take up to 18 months
Confidentiality	Patient and medical confidentiality
Geographic Specificity	Patient address, geocoded to either residential address, block group, through latitude/longitude matching, or using zip plus 4 centroid

Data Collection Process: Hospitals have six months to submit tumor registry information to one of eight regional registries. Hospitals use CNExT, or a similar application to electronically submit transactions to the registries. Five of the registries use EUREKA as their primary application, while three (Los Angeles, Orange County, and San Diego) regions have developed their own systems to capture the information, later submitting to the State's EUREKA system. In some cases, the information is identified in a manual review of Pathology laboratory records. The regional registries may also access other State agencies and departments (Vital Statistics, DMV, Voter registration) to update the status (e.g., death) of the patient, if necessary.

QC/Reliability: The EUREKA system and other registry systems have installed a version of the CDC EDITS application to perform data validation checks. Once the regional registries have received the data from hospitals, a visual check of the data are completed. The CCR then completes a 10% sampling audit of records from the registries. The multi-tiered quality review results in highly reliable data and is considered the 'best in the United States'.

System support and maintenance: The technical staff of the Public Health Institute maintains the system.

Data transfer capability: Currently, the CCR staff responds to requests for reports and files. There is no external access to the data.

Planned Enhancements: The EUREKA system has been implemented for less than a year. There are no planned enhancements at this time, but discussions are underway to consider enhancing edit capabilities, and providing access to "non-research ready" data.

Appendix B System Summaries: EUREKA (California Cancer Registry)

IT Principals:	Dr. Bill Wright, CCR Director; Steve Fuchslin, Systems Support Manager
Data dictionary:	Yes

Areas for Additional Assessment and Discussion:

- Identify feasibility of, and resources for collecting family residential history.
- Discuss improvements in release of official data.
- Identify possible methods of ongoing automated data exchange in the future.
- Determine whether GIS web/application services provided by EHTN or CDC would be desired and could be incorporated into existing reporting application architecture.
- Discuss streamlining IRB and planned enhancements in the area of data exchange and GIS (real-time geocoding and address validation, analysis and visualization applications).

Appendix B System Summaries: Medical Care Statistics Section System

Description and Use: The Medical Care Statistics Section (MCSS) provides support to the California Department of Health Services, other California agencies and interested parties by developing reports and files on Medi-Cal enrollments and fee-for-service (FFS) utilization. The MCSS is the official "point of release" for Medi-Cal data. The MCSS utilizes claims and eligibility data from the Medi-Cal Fiscal Intermediary (FI), the Medi-Cal Eligibility Data System (MEDS) and non-FI sources of information (e.g., County Operated Health Systems, Denti-Cal, and Department of Mental Health).

Paid claims data and eligibility files are created on a monthly basis. The data are used to create annual statistical reports, support fiscal forecasting activities, and provide information to a variety of users for policy, administrative and research purposes. When using these Medi-Cal data sources for preparation of data files for outside researchers, the HIPAA privacy rules are complied with. Under these rules, three levels of data sets can be provided:

- De-Identified data set – no restrictions (HIPAA or otherwise) for releasing this data. No patient identifiable information (name, address, city, county, SSN/CIN) is included, but could include age (up to age 89), sex and ethnicity.
- Limited data set – restricted use for research and must have an agreement between requestor and MCSS. Data excludes patient name, street address, SSN/CIN, but could include city and county.
- Confidential data set – restricted use and must have an agreement with MCSS and approval from the Committee for the Protection of Human Subjects. Includes claims (except First Data Bank Smartkey and rate information negotiated with the California Medical Assistance Commission), patient and demographic information.

Whenever data are requested, only the minimum data required to meet the needs of the requestor is provided (not the full dataset), and the request must be justified. More complete information on HIPAA is available on the MCSS website.

Information collected from the full Medi-Cal database, and available in one or more of the three accessible data sets includes:

• Patient demographic information	• Patient address information	• Eligibility information
• Provider information	• Claims processing information	• Diagnosis codes
• Procedure codes	• Billed and paid amounts	• Dates of service

This combined database is a very large database, collecting millions of unique claim records a year. Some data are available from about 1986. However, in 1994, an expanded claims record was implemented to capture additional data elements. Managed Care encounter data are available, but are generally not utilized, especially due to completeness issues. The results of an analysis of encounter data for CY1999 can be found on the MCSS web site at:
http://www.dhs.ca.gov/admin/ffdmb/mcss/PublishedReports/Encounter%20Data/encounter%20data.htm

Age	Originally created in 1970's, with modifications since then
Size	50 gb monthly from FFS FI paid claims 30 gb monthly from non-Medi-Cal FI sources Combined claims number over 150,000,000 per year 6 gb monthly from MEDS
Reporting Frequency	Some claims available from 1986, current format from 1994
Data Currency	Data received monthly and official within one month of receipt
Confidentiality	Patient and medical confidentiality
Geographic Specificity	Patient address available in the confidential data set

Data Collection Process: Data are collected from source systems on a monthly basis. The claims information is submitted in a standard format (e.g., '35-File' Paid Claims File format). An extract of MEDS is also received. The eligibility information can be matched with claims information using the beneficiary SSN. The MEDS extract files are archived semi-annually to serve as an historical record of eligibility.

Appendix B System Summaries: Medical Care Statistics Section System

QC/Reliability: The data are received from source systems, where internal quality review/assurance processes have been completed.
System support and maintenance: All files are maintained on the HHSDC mainframe. Some files are downloaded and maintained for use on PCs.

Data transfer capability: Currently, the MCSS staff respond to requests for reports and files. There is no external access to the source data.

Planned Enhancements: There are no planned enhancements to the MCSS database.

IT Principals:	Jim Klein, Research Specialist, MCSS
Data dictionary:	Yes

Areas for Additional Assessment and Discussion:

- Identify possible methods of ongoing automated data exchange in the future.
- Determine whether GIS web/application services provided by EHTN or CDC would be desired and incorporated into existing reporting application architecture.

Appendix B System Summaries: Patient Discharge Database

Description and Use: The Health Data and Advisory Council Consolidation Act within the California Health and Safety Code, Sections 128675, et seq. requires that every organization that operates, conducts, owns, or maintains a health facility will create and file with the Office of Statewide Health Planning and Development (OSHPD), twice annually, a report of Hospital Discharge Abstract Data Records. The hospital or their designated agent provides data for each individual in-patient discharge. The data record captures data detailed in the California Health and Safety Code, Sections 128675, et seq. Data elements include:

• Patient identification information (SSN)	• Source and type of admission	• Principal external cause of injury
• Hospital and patient's residence zip code	• Admission and discharge date	• Date for procedures
• Principal and secondary diagnoses	• Principal and secondary procedures	

The Legislature, hospitals, consultants, colleges, and universities use the Patient Discharge Database data (e.g., to determine market share, revisit strategic plans, identify new products, analyze aggregate billing patterns). OSHPD provides data to the public, health care industry, media, research community, and other State agencies, some through formal agreements that may be renewed annually (e.g., provide information to link birth and death certificates with the discharge database information). OSHPD will prepare ad hoc reports upon request for a fee. Internal health services staff can access the Patient Discharge Database standard tables, profiles and pivot tables via the Internet.

The Patient Discharge Database does not include data on outpatient doctor and emergency department visits, medical care facilities, Veteran's Administration hospitals, military base hospitals, prison hospitals, and tribal health facilities.

Age	Database in use since 1983
Size	3.8 million records added annually
Reporting Frequency	480 facilities/hospitals report every six months
Data Currency	January to June 2003 data should be available as official in early November, 2003
Confidentiality	Restrict public access to the dataset due to identifiers such as social security, date of birth
Geographic Specificity	Hospital and patient's residence zip code; patient identifiers (SSN)

Data Collection Process: Hospitals submit data, which are validated through several levels of editing in the system. Hospitals are notified electronically if they are in compliance or if corrections are needed. Hospitals must submit a report within 3 months of the close of the six-month reporting period or request an extension. Hospitals may modify their data during the reporting period. OSHPD then has 15 days to reject or approve the submittal and a second 15 days to make the data available to the public. Hospitals receive a Data Distribution Report and a Hospital Inpatient Profile. The data are also available on the OSHPD website.

QC/Reliability: Patient Discharge Database error tolerance levels are set in regulation; less than 2% of all records may have one or more errors. Quality control is performed using layered editing programs to include format (e.g., invalid entry, blank fields), relational edits (e.g., gender does not match diagnosis), readmission edits (e.g., match source to previous admission), and coding edits (e.g., diagnoses and procedures review).

System Support and Maintenance: OSHPD analysts are assigned and manage a patient discharge record caseload to include data quality control. In addition, OSHPD makes available Patient Discharge Data expertise to an Internal Resource Center that provides technical support to internal and external data users.

Data Transfer Capability: The Patient Discharge database is an Oracle database housed at HHSDC. A data warehouse merges Patient Discharge database information with the License File System to provide access to OSHPD staff. Hospital reporting, notifications, and file transfer are completed electronically.

Planned Enhancements: Senate Bill (SB) 1973 requires the expansion of data collection to include patient discharge data for hospital emergency departments, and hospital and freestanding ambulatory surgery clinics (see http://info.sen.ca.gov/pub/97-98/bill/sen/sb_1951-2000/sb_1973_bill_19980922_chaptered.html). These data will be captured within a new database as part of the Emergency Department/Ambulatory

Surgical Centers (ED/ASC) project. In addition, the primary input source will change from the OSHPD proprietary dataset to the ANSI 837 format.

IT Principals:	Candace Diamond, Manager Patient Discharge Section; Mike Kassis, CIO; Deborah Holstien, IT Project Manager, ISS Discharge Data and Accounts; Technical Improvement Committee
Data Dictionary:	Yes

Areas for Additional Assessment and Discussion:

- Discuss future collection of address data and secure access via Internet.
- Determine whether GIS web/application services provided by EHTN or CDC would be desired and incorporated into existing reporting application architecture.

Appendix B System Summaries: Response & Surveillance System for Childhood Lead Exposure

Description and Use: The Childhood Lead Poisoning Prevention Act (California Health & Safety Code 105275 to 105310) established the Childhood Lead Poisoning Prevention Branch (CLPPB) within the California Department of Health Services and requires them to compile information, identify target areas, and analyze information to design and implement a program of medical follow-up and environmental abatement to reduce childhood lead exposure. The Response and Surveillance System for Childhood Lead Exposure (RASSCLE) system was developed to capture and maintain information related to childhood lead poisoning. There is a Statewide RASSCLE system maintained within the CLPPB, and the local Childhood Lead Poisoning Prevention Programs (CLPPPs) also have the system installed to support their case management and operational activities.

As of January 1, 2003, laboratories are required to report all blood lead levels (BLL). Prior to 2003, laboratories were required to report only elevated BLLs. The laboratories complete a Laboratory Reporting Form (LRF) and notify the State of BLLs exceeding a specific threshold. The LRFs may be submitted electronically, or as a paper form. The information from the LRFs is entered into the State RASSCLE system, and copies transmitted to the local health departments for follow-up. By 2005, LRFs must be submitted electronically.

Data elements collected include:

• Patient identification information	• Patient address	• Laboratory information
• Ordering physician identification and information	• Date of blood draw	• Date of laboratory analysis
• Test result and blood lead level	• Investigative results	• Test type (i.e., venous or capillary sample)

Data are summarized and reported to the CDC on a quarterly and annual basis. The Childhood Lead Poisoning Prevention Branch also provides epidemiological research support and responds to external requests for data.

Age	1992; Age of data in system is _____
Size	Number of records _____
Period reported	Daily
Data Currency	Data currency is dependent upon the lead level report priority for data entry as described in the Public Health Nursing Manual – Period for official data to rate against criteria is _____
Confidentiality	Restrict public access to the dataset due to the maintenance of patient information
Geographic Specificity	Address, census tract, zip code

Data Collection Process: On a daily basis, laboratories transmit electronic or hard copy Lab Reporting Forms that detail lead level test results. The State enters lead level reports into the State RASSCLE database and exports the data to five CLPPPs electronically. The remainder of the CLPPPs receive lead reports manually. Among those BLLs that are not reported to the CLPPP electronically, elevated BLLs and BLLs associated with existing cases are prioritized highest for data entry into the REASSCLE system. Data are maintained at both the State and local jurisdiction level. The State requires that the CLPPPs investigate BLLs meeting case definition. Local jurisdictions investigation of cases may result in the completion of a Lead Poisoning Follow-Up Form; forms for high lead levels are forwarded to the State.

QC/Reliability: Quality Control checks include reconciling State and local database information. There are system checks upon data entry for duplicate records or record information. In addition, the Branch will generate audit reports to identify missing or follow-up reports and notify the appropriate local health departments to correct the record.

System support and maintenance: One FTE Database Programmer supports RASSCLE. One system specialist provides technical support and system testing, and one research associate provides CLPPP training, support and development.

Appendix B System Summaries: Response & Surveillance System for Childhood Lead Exposure

Data transfer capability: The Childhood Lead Poisoning Prevention Branch does not provide external access to the FoxPro database. Data are summarized within reports.

Planned Enhancements: The Childhood Lead Poisoning Prevention Branch is currently working on a replacement database for RASSCLE. This system, RASSCLE2, will be a single database, web-based system to provide State and local jurisdictions with a single point for data entry. The Branch anticipates implementing the system during CY 2004. The CLPPB is completing work on an Electronic Laboratory Reporting (ELR) system. The ELR will allow laboratories to report BLLs to the CLPPB electronically, and ensure capture of reported BLLs.

IT Principals:	Jeff Sanchez, Section Chief, Health Information Assistance Section
Data dictionary:	Yes

Areas for Additional Assessment and Discussion:
None

Appendix B System Summaries: SENSOR Asthma Database

Description and Use: The SENSOR Asthma Database is a federally funded component of the Sentinel Event Notification System for Occupational Risks (SENSOR). This database captures data on adult work-related asthma, including the cause of the illness. The California SENSOR Asthma Database is considered the most representative work-related asthma surveillance database in the nation due to the statewide administrative-based reporting process. Data captured include:

• Demographic and administrative data	• Dates types and route of exposure	• Lost time
• Industry and occupation data	• Physician identification	• Content of exposure (agent information)

The SENSOR Asthma Database data are used at several levels for prevention. Educational materials are provided to each reported patient as well as to employers in a specific industry with potential risk. DHS staff work cooperatively with, and provide technical assistance to, employers to improve safety in the workplace. Data are also used to identify risk factors and develop prevention activities. In addition, data are used to inform the development of standards, regulations and law. The SENSOR asthma program submits de-identified raw data annually to CDC.

The SENSOR Asthma Database currently resides on two databases. Initial reports are entered into an Epi Info database and interview/follow-up data are entered into a FoxPro database. DHS staff are currently developing an Access database to replace these two system, with an estimated October 2003 deployment timeline.

The database may not capture all work-related asthma cases. It does not capture information for individuals who fail to visit the doctor, for cases where the doctor fails to recognize the illness or categorize it as work-related, and for cases where the workers' compensation insurance company fails to forward a doctor's report. The database also does not routinely receive reports for the self-employed, maritime workers, and Federal employees.

Age	Developed in 1987
Size	3,100 records; 300 records added annually
Reporting Frequency	Daily
Data Currency	Official data are available 6 months to one year after the reporting date
Confidentiality	Do not share patient identification information and always maintain strict confidentiality for interview information per the requirements of our Human Subjects approval. HIPAA has only reinforced this. HIPAA has made it more difficult to retrieve medical records.
Geographic Specificity	Employee, Employer, and Doctor street address to include zip and Federal Information Processing Standards (FIPS county and state) code

Data Collection Process: Physicians submit a Doctor's First Report (DFR) for patients with a work-related injury to the appropriate workers' compensation insurance company. The insurance company is required by law to forward the reports to the State Department of Industrial Relations (DIR). Physicians are compensated for treatment of these patients via these reports.

DIR forwards the reports to the DHS, Occupational Health Branch. DHS staff sort the reports by type of illness or injury. For reports of an asthma incident, DHS staff contact the patient to confirm the illness, identify the type of asthma, and conduct a 20 minute telephone interview. If the respondent is unreachable or refuses, DHS staff request medical record information from the medical care provider. DHS estimates that they may be receiving only one third of reports for all incidents. All cases are entered into the SENSOR Asthma Database, including variables indicating if cases are confirmed and classified. As part of the patient's consent to the interview, personal information is collected, but not available outside of the database.

QC/Reliability: The SENSOR Asthma Database includes a very reliable but incomplete data set (e.g., doctor's are not asking the right questions, data often do not include the content of exposure). Reported data are entered into the database and data entry is reviewed for accuracy and to avoid record duplication.

System Support and Maintenance: 2.25 FTE technical staff support and maintain the system.

Appendix B System Summaries: SENSOR Asthma Database

Data Transfer Capability: The SENSOR Asthma Database staff have access to the data.

Planned Enhancements: DHS plans to evaluate workers' compensation electronic reporting of the employer's first reports via the California Workers' Compensation Information System (WCIS). The WCIS uses Electronic Data Interchange (EDI) to collect comprehensive information from claims administrators to help the DIR oversee the state's workers' compensation system.

IT Principals:	Jennifer Flattery, Research Scientist, DHS Occupational Health Branch
Data Dictionary:	Yes

Areas for Additional Assessment and Discussion:
None

Appendix B System Summaries: SENSOR Pesticide Illness Database

Description and Use: The SENSOR Pesticide Illness Database is a federally funded component of the Sentinel Event Notification System for Occupational Risks (SENSOR). This database captures data on work-related, acute pesticide illnesses including the cause of the illness. Data captured includes:

• Administrative and demographic data	• Type of exposure	• Agent information
• Industry and occupation data	• Route of exposure	• Medical diagnosis (this is not as important to pesticide case classification, but is captured)
• Signs and symptoms	• Type care and lost work time	• Enforcement Agency findings

System users include the National Institute for Occupational Safety and Health (NIOSH) that collects an annual data set, with no personal identifiers, from five SENSOR states to analyze aggregated data. In addition, data are used to define the signs and symptoms related to acute pesticide illness, effect pesticide label changes to reduce risk, support draft legislation to improve the regulation of pesticides, and document primary health effects.

The SENSOR Pesticide Illness Database excludes non-occupational cases, and illness related to disinfectant exposures.

Age	Access database created in 1997
Size	2,000 cases with 400 cases added annually
Reporting Frequency	Daily
Data Currency	Records will take up to a year from report date to process. Last dataset made available to NIOSH was from 1998-2001
Confidentiality	Following California's Health and Safety Code, we protect the identity of the individual. We also maintain strict confidentiality for our worker interviews per the requirements of our Human Subjects approval.
Geographic Specificity	Employee address, injury site and treatment site to include zip and Federal Information Processing Standards (FIPS) code

Data Collection Process: Physicians submit a Doctor's First Report (DFR) for patients with a work-related injury to the appropriate workers' compensation insurance company. The insurance company is required by law to forward the reports to the State Department of Industrial Relations (DIR). Physicians are compensated for treatment of these patients via these reports.

DIR forwards the reports to the DHS, Occupational Health Branch. DHS staff sort the reports by type of illness or injury. DHS staff will stamp each report with a unique identifier and enter each report into the SENSOR Pesticide Illness database. After which, DHS staff will request medical records for each work-related, acute pesticide illness report from the medical provider. In addition, DHS staff will add additional report sources of information to enhance case ascertainment (e.g., Pesticide Incident Report from the Department of Labor and Statistics, Pesticide Episode Transmittal Records from the Department of Pesticide Regulation) to the uniquely identified report using variables: such as name, social security number, date and circumstances of exposure. Once information has been compiled on the record, it is abstracted and case classified (assigned a status).

DHS will select reports meeting the criteria for targeted field investigations. Reports are investigated based on time and staff availability. For these reports, DHS staff will conduct follow-up such as interviews and site visits. All findings related to the standardized variables are integrated into the records and used for case classification.

QC/Reliability: DHS staff perform a manual check for quality of record abstraction and data entry, check for duplicate records, assign event identification and perform a series of queries to link records to the same event. In addition, staff use Statistical Analysis Software (SAS) to analyze data for outliers.

System Support and Maintenance: Two full-time staff support and maintain the database.

Data Transfer Capability: The SENSOR Pesticide Illness Database staff have access to the data.

Appendix B System Summaries: SENSOR Pesticide Illness Database

Planned Enhancements: Inclusion of workers' compensation electronic reporting of employers' first reports via California Workers Compensation Information System (WCIS).

IT Principals:	Ximena Vergara, Research Associate and John Beckman, Research Associate, DHS Occupational Health Branch
Data Dictionary:	Yes

Areas for Additional Assessment and Discussion:
None

Appendix B System Summaries: Automated Vital Statistics System (AVSS)

Description and Use: The Center for Health Statistics maintains vital records. The Automated Vital Statistics System (AVSS) is used to produce birth certificates and electronically transmit this information from hospitals and birthing centers to the local health department and subsequently to the State Registrar. Some local health departments also key enter selected death certificate data into AVSS, and a few local jurisdictions are using AVSS on a pilot basis to produce death certificates and electronically transmit information from funeral establishments to the local health department. For additional information see http://www.avss.ucsb.edu/history.htm. Data captured includes:

• Family or individual demographic information	• Complications of delivery	• Survivor information
• Conditions contributing to death	• Mother's history of births	• Attending physician or coroner diagnoses
• Location of event	• Medical contact information	• Laboratory tests and results

The State Registrar produces an annual validated statistical master file, usually available 12-18 months after events are recorded. This statistical master file is the primary source for the reporting of vital statistics. Counties use the master file to extract data and generate reports. AVSS also contains a set of standard reports as well as a report generator that is used to query the local AVSS database. In addition, the State Registrar produces annual reports, which include aggregated data and statistical trends. Other data users include private researchers, universities, federal government agencies, the Department of Motor Vehicles, the Social Security Administration, and various social services organizations.

Recent legislation (i.e., SB247, SB1614) impacts the availability of vital statistics data. Key personal identification fields (e.g., social security number, mother's maiden name) are considered protected. Information requests redact or mask this information. Confidential data are released only upon approval from the Committee for the Protection of Human Subjects and for a critical need such as fraud investigations, legislative intervention, Department of Justice investigations, and research.

Age	Database created in 1980
Size	Estimated 240,000 death records and 530,000 birth records per year
Reporting Frequency	Daily
Data Currency	Data forwarded from the County is considered official upon receipt by the State; last issue CY 2002 master file
Confidentiality	Restrict access to personal identification information
Geographic Specificity	Captures zip code, census place code, county of residency. AVSS provides automated lookup of street addresses for geo-coding at the census tract level.

For additional information on AVSS specifications, see http://www.avss.ucsb.edu/facts.htm.

Data Collection Process: AVSS data are captured through the electronic transmission of birth and death data or through data entered from hard copy forms. The State has experimented with paying counties $1.00 for every death record keyed into AVSS. The State enters data on behalf of several smaller counties. Users can enter, validate and correct birth records online prior to registration. State staff review and validate data.

QC/Reliability: AVSS includes 5 to 6 years of reliable data. All AVSS data entry is subjected to error checks, with errors ranging from warnings to the complete rejection of incorrect values. Validation procedures are performed to compare values from different data fields for consistency. Vital event records in error or that need to be updated with additional or changed information are amended using hard copy attachments to the original certificate. AVSS is updated to reflect amendments that are registered within a year of the birth. Audit trails are maintained on all user interactions with records.

Death records may take weeks, months or years to complete (e.g., cases may remain open during coroner investigations). In addition, vital records may include incorrect personal and demographic data for individuals who do not want to be located or are transient.

System Support and Maintenance: State staff (3 FTE) and UC Santa Barbara staff (4 FTE) maintain the system and provide Help Desk support to users.

Appendix B System Summaries: Automated Vital Statistics System (AVSS)

Data Transfer Capability: In addition to transmitting data by means of telecommunications, AVSS can also use electronic media such as tape or diskette. Electronic records can be communicated between facilities by specifying date or file number ranges. Tapes or diskettes meeting the California Birth Certificate (CBC) Electronic Data Submission Requirements can be easily created. AVSS data can be exported in ASCII format to other computer systems by means of magnetic tape or diskette.

The ability to import data into AVSS from non-AVSS computers is operational in several sites. AVSS imports birth records from the Site of Care and Southern California Kaiser hospital systems.

Planned Enhancements: The State Registrar plans to expand the number of sites accessing AVSS through the Internet and to expand the number of death entries. An Electronic Death Registration System (EDRS) is also in planned development as a web based Internet application for use by funeral directors, physicians, coroners, medical examiners and health departments.

IT Principals:	Mike Quinn, Chief, DHS Office of Health Information and Research, Vital Statistics Section, Mike Rodrian, State Registrar and Chief, DHS Health Information and Strategic Planning, Center for Health Statistics
Data Dictionary:	Yes

Areas for Additional Assessment and Discussion:

- Identify possible methods of ongoing automated data exchange in the future.
- Determine whether GIS web/application services provided by EHTN or CDC would be desired and incorporated into existing reporting application architecture.
- Discuss streamlining IRB and planned enhancements in the area of data exchange and GIS (real-time geocoding and address validation, analysis and visualization applications).
- Clarify whether family or individual demographic information includes mother's/father's occupation/industry.