Sources of Data & Crash Data Management

Fall 2021

Sources of Data

- Traditional Data
  - Crash Data (Police reports, self-reporting, etc.)
  - Traffic Conflicts/Surrogate Measures of Safety
  - Hospital
  - Crash Reconstruction
  - Insurance (Private)
- Naturalistic Data
  - Instrumented Vehicles
  - Vehicle Blackboxes
  - Safety Systems (Onstar, etc.)
- Social Media/Crowd Sourcing Data
  - Twitter, Facebook, Instagram, Waze
- Disruptive Technological Data (enhance or new technology)
  - Cellphones
  - Video Recording/Processing
  - Roadside Sensors (Toll Roads, etc.)

Links are provided at the end.

Traditional Data

- General Characteristics
  - Despite the flaws, crash data provide the ultimate sources of information about the safety performance of roadway elements.
  - Unfortunately, people get hurt or killed and properties get damaged before finding this performance.
  - Highly dependent on severity of the crash
- Crash data used for:
  - Identification of hazardous sites
  - Benefit-costs analysis
  - Safety relationships (statistical models)
  - Highway design process
  - Policy development

Links are provided at the end.

Table 2.4 Important Variables Collected from Crash Data

Variable	Description
Identification Number	Each crash report should have its own identification number. This ensures that each crash is unique and can be easily traceable.
Location	The location can be identified using a linear system, such as control-section mile point on a predefined maps maintained by the transportation agency. More recently, most agencies are now reliably coding crash data using the geographic information system (GIS) technology.
Date and Time	These two variables can be used to assign crashes for different seasons and whether the crash occurred during nighttime, dusk, dawn or daytime conditions.
Severity	This is used to characterize the most severe injuries among all the occupants or vulnerable road users (pedestrians or bicyclists). For example, if a crash has three injures, one incapacitated (Type A) and two possible injuries (Type C), the crash will be classified as incapacitating injury (Type A).
Collision Type or Manner of Collision	This variable describes the characteristic of the crash, such as right-angle, side-swipe or left-turn/through collision.
Direction of Travel	This variable explains the direction or trajectory of each vehicle or road user involved in the crash.
Alcohol or Drugs	This variable explains if any of the drivers or vulnerable road users was under the influence of alcohol or drugs. This variable will be often be updated in the report after the crash to account for the time needed to get laboratory results back.
Vehicle Occupants	This variable describes the gender and age of each vehicle occupant or road user. It may include the legal driving and insurance statuses of drivers.
Vehicles Involved	This one describes the characteristics of each vehicle. This variable defines the crash as being a single-vehicle or multivehicle event.
Narratives	This section of the report is usually not coded electronically. However, the narrative is very important since it provides information about the crash process (based on the testimony of witnesses and the visual assessment of the officer). It is usually accompanied by one or more figures or sketches which helps explain what happened. Based on the authors' experience, many research projects involved the review of these narratives for validating the electronic databases. This is a very time consuming and costly process.

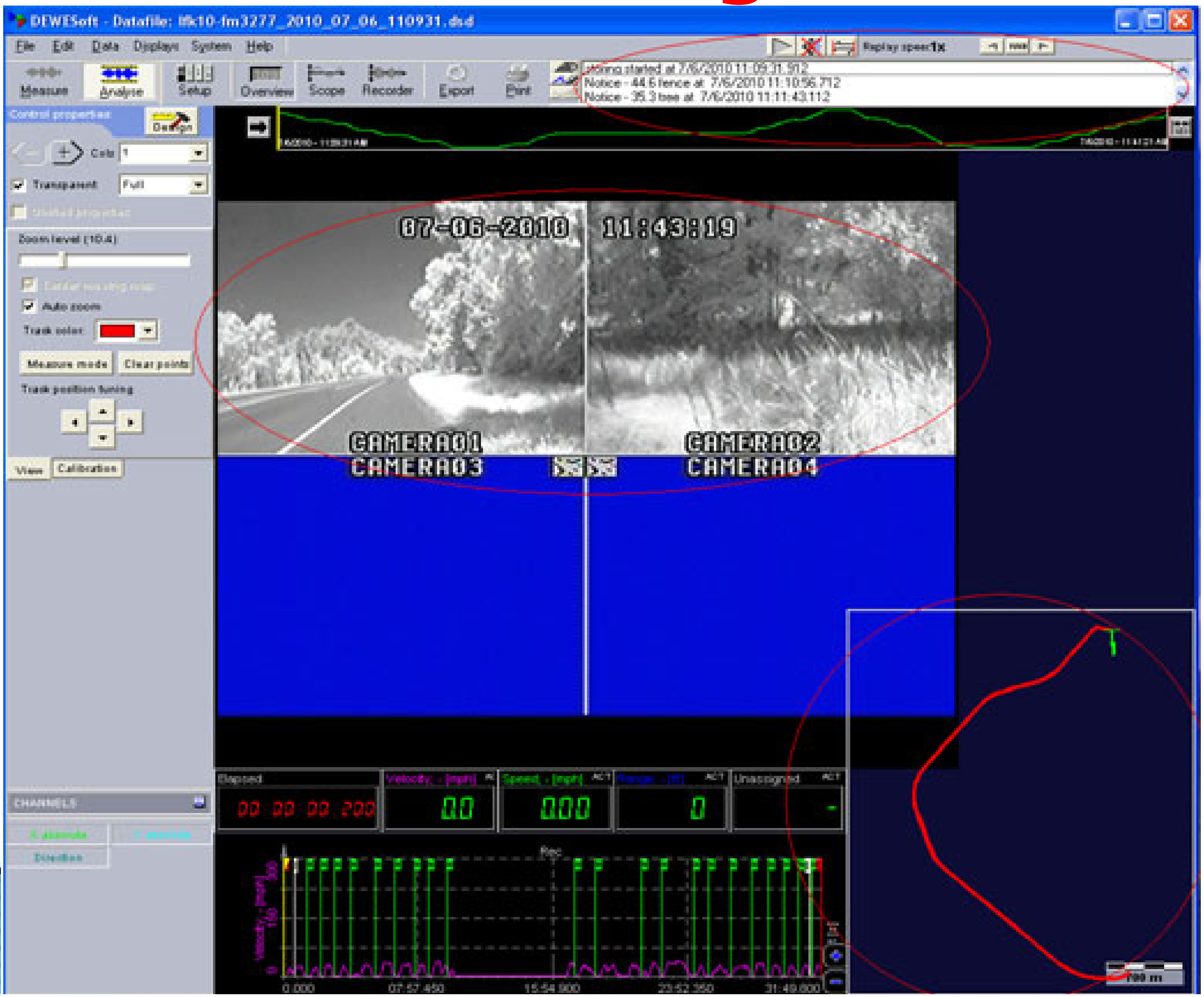
Table 2.5 Common Variables Found in Roadway Data

Location (control-section mile point or GIS) Segment Length Type of Pavement Traffic Control at Intersections Speed Limit Road Alignment (tangent, curve) Road Surface Condition Right-of-Way Width Parking	Highway Classification (Freeway, Arterials, etc.) Type of Lane and Width Type of Shoulder and Width Number of lanes Divided/Undivided Lighting

Table 2.6 Traffic Flow Data

Location (control-section mile point or GIS)
Annual Average Daily Traffic/AADT (vehicles/day)
Average Daily Traffic/ADT (vehicles/day)
Traffic Mix (heavy vehicles, motorcycles, passenger cars, etc.)
Speed Distribution
Short Counts (hourly volumes, 15-minute values, etc.)
Vehicle Occupancy (on instrumented urban freeway segments)
Traffic Density (on instrumented urban freeway segments)
Turning Movements at Intersections

Video Logs



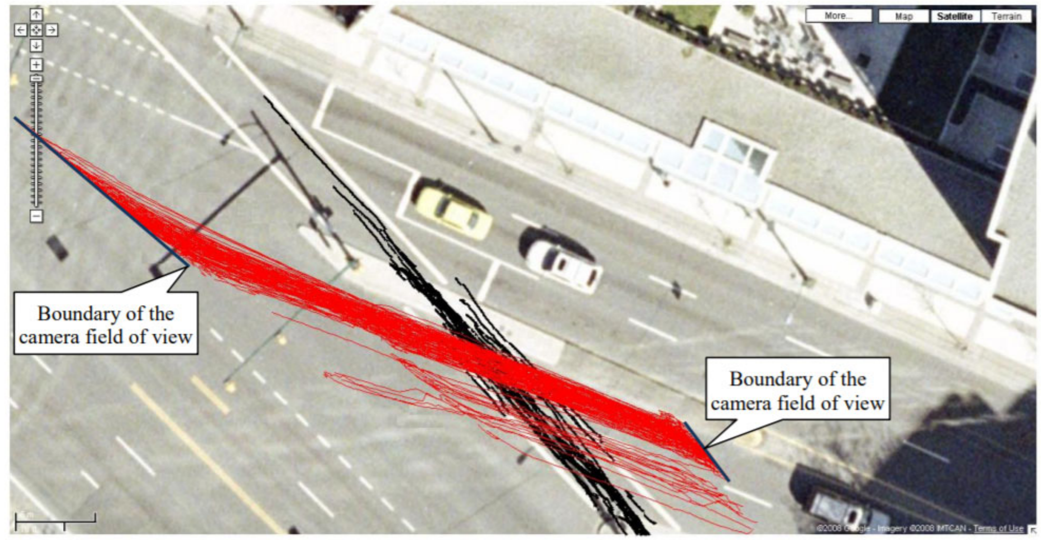
DEWESoft - Datafile: tfk10-fm3277_2010_07_06_110931.dsd
File Edit Data Displays System Help
Measure Analyse Setup Overview Scope Recorder Export Print
Replay speed 1x
Notice - 44.6 fence at 7/6/2010 11:10:56.712
Notice - 35.3 tree at 7/6/2010 11:11:43.112
Transparent
Full
Zoom level (10.4)
Auto zoom
Track color
Measure mode
Clear points
Track position tuning
View
Calibration
07-06-2010
11:43:19
CAMERA01
CAMERA02
CAMERA03
CAMERA04
CHANNELS
Direction
Elapsed
Velocity - [mph]
Speed - [mph]
Unassigned
0.0
0.00
0
0.000
07:57.450
15:54.900
23:52.350
31:49.800

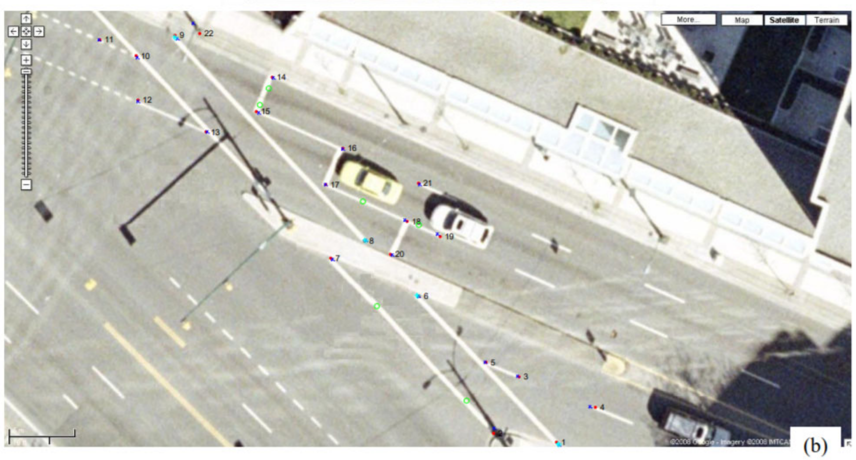
Google Earth

StreetView

Video Processing

Video Processing





Automated Analysis of Pedestrian-Vehicle Conflicts Using Video Data. Ismail et al. (2009) https://doi.org/10.3141/2140-05

Video Processing

Automated Analysis of Pedestrian-Vehicle Conflicts Using Video Data. Ismail et al. (2009) https://doi.org/10.3141/2140-05

Data Users

- Traffic Engineers: elimination of hazardous sites; highway design, etc.
- Police force: enforcement location, etc.
- Researchers: understanding the crash process; safety relationships, etc.
- Decision-makers: alcohol measures, speed limit
- Prosecutors: transportation-related criminal lawsuits, eye witness statements, etc.
- Insurance companies: set premiums, types of vehicles, age of drivers,
- Vehicle manufacturers: research for safer vehicles

Crash Report

- Quality of data highly dependent on the officer at the scene of the crash
- Most important data are collected for potential criminal prosecution; other data less important
- Sometimes officers fill out the report on the scene while others do it at the end of the day (see next overheads)
- Important to have an open line of communication between the engineering and police departments
- Now, crash reports and process for collecting all electronic

Critical Information

- Geographical location (intersections, mile-point, GIS)
- Date (year, day of week, time of day, etc.)
- Type of involvement (vehicle, driver, occupants, etc.)
- Outcome (severity)
- Environmental conditions (weather, lighting, road surface conditions)
- Characteristics of collision (direction of road users, errors, collision type)

Critical Information

Table 4-1 Accident data – Recommended information

ACCIDENT IDENTIFICATION NUMBER	LOCATION
- A unique number, which prevents accident data from being entered twice (should be combined with police station reference number).	- The exact road location where the accident occurred. It can be described by: - narratives; - X, Y coordinates of a uniform coordinate system; - highway number and kilometer post rounded off, for example, to the nearest 100 m; - distance from node; - distance from known point. ***Accident location methods***
DATE AND TIME	**COLLISION TYPE**
- The exact date (preferably with four digits for the year, two digits for the month and two digits for the day). - Although the day of the week can be computed from the date, it may be useful to have a separate data item to store this piece of information. This is especially true in the case of a non-computerized database.	- By one or some combinations of the followings: - narrative description; - sketch; - code.
VEHICLE	**VEHICLE MANOEUVRES AND SKETCH**
- For each involved vehicle: - type; - make; - year of manufacture; - vehicle identification number (VIN). - The VIN is a multi-character identification number placed either on the body of the vehicle or included on the vehicle registration certificate. Through the VIN, a number of variables, such as type, model, year of manufacture, body type, engine size, restraint system, etc., can be easily accessed in the vehicle registration file.	- The description codes of the manoeuvre of each vehicle/ participant or an accident sketch. - To be useful, this sketch must include the following information: - each vehicle's direction and manoeuvre; - vehicle identification (veh1, veh2, etc.); - reference points; - important measures; - scale.
[illegible]	[illegible]

Taken from the 200[illegible]rsion of the PIARC’s Road Safety Manual

Critical Information

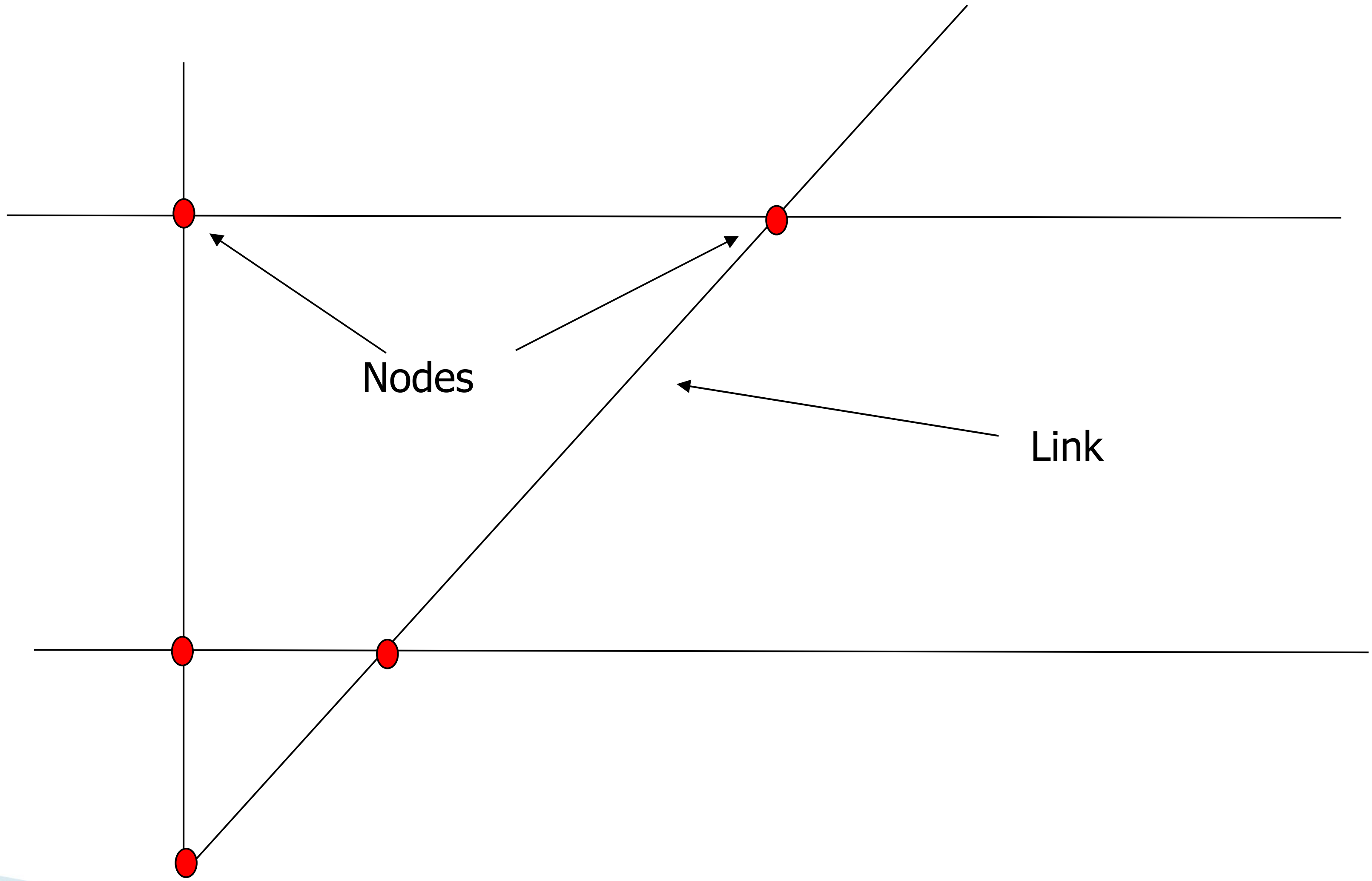
DIRECTION OF TRAVEL	**CONTRIBUTING CIRCUMSTANCES**
- The direction of each participant, for example: - from A to B; - the direction of increasing/decreasing mileage.	- What were the environmental conditions such as: - whether (rainy, windy, foggy, etc.); - road surface conditions (icy, wet, debrid on the road, etc.); - other contributing factor.
CASUALITIES	**DRIVER / PASSENGER**
- By severity of injury for each casualty, for example: - fatal; - serious injury; - light injury. - The injury severity, determined by a police officer, is often largely subjective and does not always reflect the injury scale used by a hospital. A more accurate injury severity number can be retrieved from hospital files.	- name, sex, age; - location in the vehicle; - driving licence number; - driving experience.
RESTRAINT USE	**ALCOHOL / DRUGS**
- seat belt, helmet, children safety seat; - whether air bag (deployed or non-deployed).	- By the result of the alcohol test: - on the spot; - in the police station or hospital.
SPEED	**EXTENT OF PROPERTY DAMAGE**
- driver and eyewitness statement; - length of skid marks on the road - tachograph of the truck.	- By estimation
NARRATIVES	
- Narratives are key elements of a report form. They often complete the picture about the accident with some useful additional information that cannot be coded. It is also the most time consuming part of the process, therefore tends to be lacking. Enough space should be provided on the report form to accommodate all information that the police officer finds important.	

Taken from the 2003 version of the PIARC's Road Safety Manual

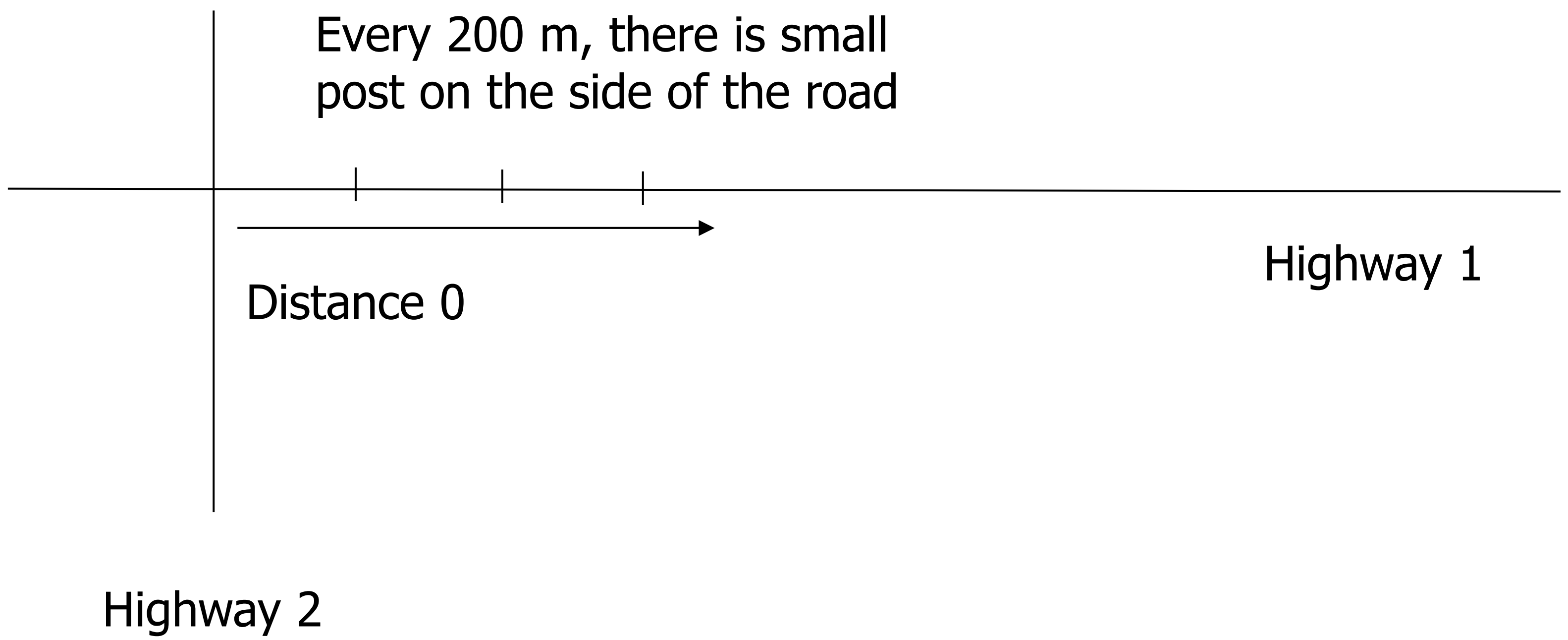
Location Methods

- Location of the crash is the most important aspect of data collection:
  - Estimate hazardous sites that experience more crashes than what would be expected
  - Provide a usually way to link different databases
- Three methods: Link-Node, Route-Km Post (aka Control-Section), GPS coordinate

Link-Node



Route-Km Post



GPS System

- Provides X and Y coordinates of a given geographical coordinate system
- Advantages: human error free; many software tools include GPS coordinate; can estimate in the Z-coordinate (depth)
- Disadvantages: Accuracy used to be an issue, but now the location pinpoint is very accurate (e.g., look at the location app on your smartphone).

Road System Inventory

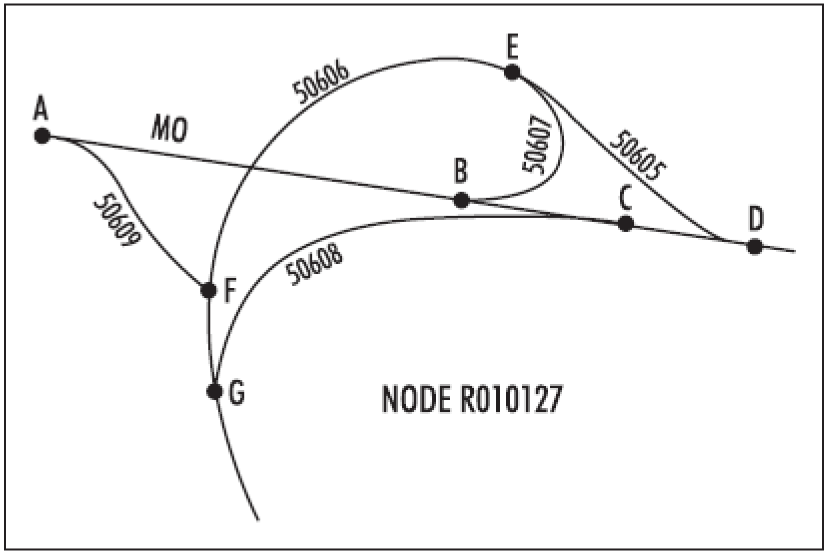
Reference map

Location accuracy depends not only on the officer's description but also on the level of detail and accuracy of the maps the officer uses to identify the location. Detailed maps are crucial when an accident occurs at an intersection with many legs or on a freeway ramp. A unique reference number may be assigned to entrance and exit ramp junction points.

The prerequisites of a successful accident location method (link-node or route-km post) are in general (Figure 4-2):

- easily recognizable reference landmarks along the road;
- a sufficiently detailed map accurately reflecting the road inventory file.

Figure 4-2 Link and node map at interchange



Links between each pair of nodes are assigned a unique number. Such maps help police officers at the scene of an accident identify the location quickly and accurately.

When roadworks have changed the length of a road and the km posts have not been altered to reflect the real distances, the posts no longer indicate the true km point of a location. Calculation algorithms must then be used to convert the reported km posts into true values.

Taken from the 2003 version of the PIARC's Road Safety Manual

Crash Data Storage

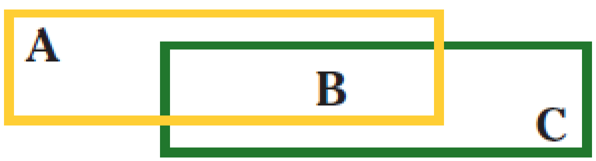
- Old way: paper trails
- All crashes are coded electronically (e.g., CRIS in Texas; FHWA's HSIS)
- Now, commercial programs available that displays visually simple crash statistics
- Internet-based visual tools (using GIS/GPS capabilities: ARCView, etc.)
- DBF, SAS, TXT files

Crash Data Limitations

- Not all crashes are reportable
  - Personal injury (definition)
  - PDO (limit, changes over time)
- Not all crashes are reported
  - Ignorance of the law
  - Victim's unawareness of injury at the time of the collision
  - Desire to avoid bureaucracy
  - Desire to avoid insurance company penalties
  - Type of collision (single vehicle versus multivehicle)
  - Type of users (bicyclists and pedestrians)
  - Police force (report level varies by jurisdiction)
- Reported crashes may contain errors
  - On report, fraudulent claims

Crash Data Limitations

Figure 4-6 Comparison of accident files (police and hospital)



A = accidents recorded by the police only
B = accidents recorded by the police and the hospital
C = accidents recorded by the hospital only

Definition 1 = $\frac{B}{B+C}$ This definition covers those hospital file cases for which a police report is found.

Definition 2 = $\frac{A+B}{B+C}$ This definition compares the total number of records found in the police file with the total number of cases in the hospital file.

Definition 3 = $\frac{A+B}{A+B+C}$ This definition, theoretically the most correct, compares the proportion of the total police records with the total number of cases in both the police and the hospital files.

Taken from the 20[illegible] version of the PIARC's Road Safety Manual

Crash Data Limitations

Figure 4-7 Accident report – The golden mean

An increase in the number of questions on a report form may lead to a sharp drop in the number of reported accidents. Some questions may not be answered by the police officer because of oversight rather than any intention not to answer.

↔

A short, one page report form may lead to higher reporting rates, but provides less detailed information. This may result in insufficient data to carry out comprehensive studies.

Taken from the 200[illegible] version of the PIARC's Road Safety Manual

Integrated Data Files

Key elements of an information system

The key elements to success in maintaining a safety information system are the following:

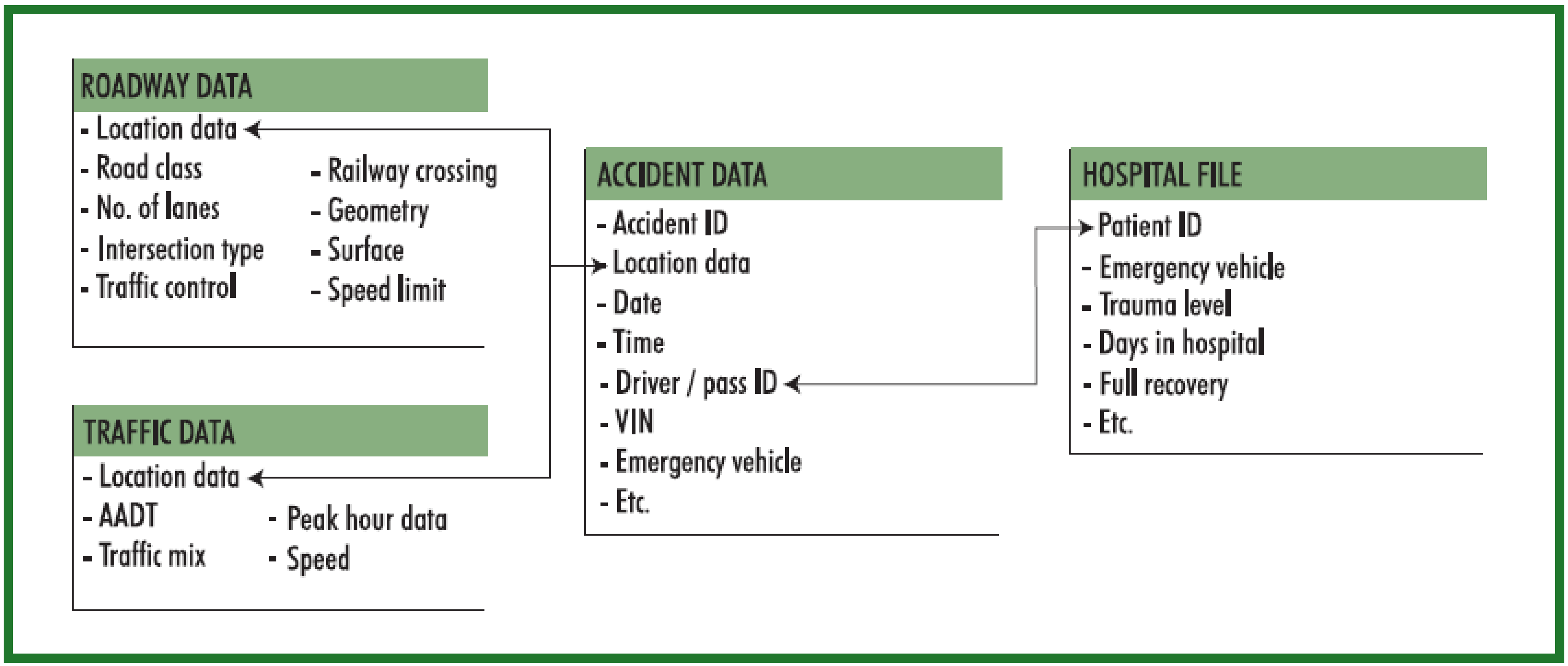
- clear definitions and terminology (this ensures that everyone knows what information a particular data item contains);
- a continuous dialogue with potential users, with changes being made to the system to meet their needs;
- fast and easy access for all user groups and levels;
- provision of timely information;
- modest centralization (data should be entered near source);
- reducing data-maintenance costs (by choosing the right program language and making sure the system is flexible enough and open to future upgrades);
- consistency in the same data item across different files (attention to data items that make linking possible).

Taken from the 2003 version of the PIARC's Road Safety Manual

Integrated Data Files

Linking Files

Figure 4-9 Linked file



Taken from the 200[illegible] version of the PIARC's Road Safety Manual

HSM Data Need

Crash Data — The data elements in a crash report describe the overall characteristics of the crash. While the specifics and level of detail of this data vary from state to state, in general, the most basic crash data consist of crash location; date and time; crash severity; collision type; and basic information about the roadway, vehicles, and people involved.

Facility Data — The roadway or intersection inventory data provide information about the physical characteristics of the crash site. The most basic roadway inventory data typically include roadway classification, number of lanes, length, and presence of medians, and shoulder width. Intersection inventories typically include road names, area type, and traffic control and lane configurations.

HSM Data Need

Traffic Volume Data — In most cases, the traffic volume data required for the methods in the HSM are annual average daily traffic (AADT). Some organizations may use ADT (average daily traffic) as precise data may not be available to determine AADT. If AADT data are unavailable, ADT can be used to estimate AADT. Other data that may be used for crash analysis includes intersection total entering vehicles (TEV), and vehicle- miles traveled (VMT) on a roadway segment, which is a measure of segment length and traffic volume. In some cases, additional volume data, such as pedestrian crossing counts or turning movement volumes, may be necessary.

Research Results Digest 329
HIGHWAY SAFETY MANUAL DATA NEEDS GUIDE (June 2008)
http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rrd_329.pdf

HSM Data Need

Limitations of observed crash data include:

- **Data quality and accuracy**
  - **Data entry**—typographic errors;
  - **Imprecise entry**—the use of general terms to describe a location;
  - **Incorrect entry** —entry of road names, road surface, level of crash severity, vehicle types, impact description, etc.;
  - **Incorrect training**—lack of training in use of collision codes;
  - **Subjectivity**—Where data collection relies on the subjective opinion of an individual, inconsistency is likely. For example, estimation of property damage thresholds or excessive speed for conditions may vary.
- **Crash reporting thresholds and the frequency-severity indeterminacy**
- **Differences in data collection methods and definitions used by jurisdictions**

Naturalistic Data (SHRP2)

What is SHRP 2

- The central goal of SHRP 2 Safety research is to address the role of driver performance and behavior in traffic safety. This includes developing an understanding of how the driver interacts with and adapts to the vehicle, traffic environment, roadway characteristics, traffic control devices, and the environment. It also includes assessing the changes in collision risk associated with each of these factors and interactions. This information will support the development of new and improved countermeasures with greater effectiveness.

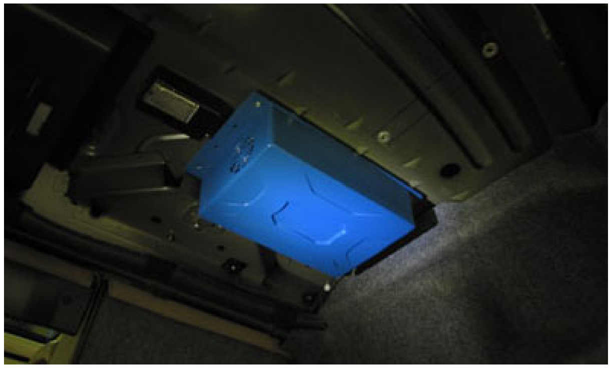
Objectives of SHRP 2

- SHRP 2 has four focus areas: Capacity, Reliability, Renewal, and Safety. The Safety focus area is conducting the NDS. It will allow us to record and study the driving behaviors of a large sample of drivers in their personal vehicles.
- http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/PublicationsSHRP2.aspx
- http://www.shrp2nds.us/
- http://www.ctre.iastate.edu/shrp2-s04a/

Equipment

The installed instrumentation or data acquisition system will collect data using a number of sensors and video cameras whenever the vehicle is running. The data acquisition system (DAS) that will be used for this study will compile data from the vehicle network as well as from sensors added for this particular study.

Head Unit



Main Unit

Radar

Locations


Seattle, WA
Buffalo, NY
State College, PA
Bloomington, IN
Durham, NC
Tampa, FL

Products of SHRP 2

- The resulting data, expected to exceed 1 petabyte in size—about the size of a million 1-gigabyte USB flash drives—will provide a wealth of information regarding driving behavior, lane departures, and intersection activities, which are anticipated to be of interest to transportation safety researchers and others for at least 20 years.
- All future research efforts that seek to use the data collected in the SHRP 2 NDS will require institutional review board (IRB) approval.
- In addition, researchers must establish a data-sharing agreement that guarantees privacy.

Studies produced from SHRP 2

- http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/Pages/The-SHRP-2-Naturalistic-Driving-Study-472.aspx#publications
- 100-Car study:
  - https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/100carmain.pdf
- Strategic Highway Research Program (SHRP2): https://insight.shrp2nds.us/ (https://vtnews.vt.edu/articles/2015/05/050115-vtti-datasharingproject.html)

Sources of Data - Traditional

Highway Safety Information System: http://www.hsisinfo.org/

Fatality Analysis Reporting System: https://www-fars.nhtsa.dot.gov/Main/index.aspx

National Automotive Sampling System: https://www.nhtsa.gov/national-automotive-sampling-system-nass/nass-general-estimates-system

General Estimates System: http://www.nhtsa.gov/NASS

Crashworthiness Data System: https://www.nhtsa.gov/national-automotive-sampling-system/crashworthiness-data-system

Crash Outcome Evaluation System: https://www.nhtsa.gov/crash-data-systems/crash-outcome-data-evaluation-system-codes

Model Minimum Uniform Crash Criteria: http://www.mmucc.us/

Bureau of Transportation Statistics: https://www.bts.gov/content/motor-vehicle-safety-data

Each State Agency should have data available.

Local cities or counties may have data available.

Sources of Data – ND/DTD

UDRIVE (Europe): https://results.udrive.eu/

UMTRI (Michigan): http://www.umtri.umich.edu/our-focus/naturalistic-driving-data

Australian Naturalistic Study: http://www.ands.unsw.edu.au/

Streetlight Data: https://www.streetlightdata.com/ (exposure)

Safe2save Data: https://safe2save.org/ (exposure)