Implementing a Distributed Responsibility Approach to Improving Data Quality

DAVID B. BROWN AND CAROLYN L. MCCREARY

An approach to improving data quality that is based on the concept of distributed responsibility is presented. This concept requires a direct distribution of data benefits throughout the data-responsible organizations. Premises on data quality are presented to formulate the problem so that the concept of distributed responsibility can be understood. Several accident data systems are discussed as examples of distributed systems, but proposals for continued development of this concept are strongly encouraged.

Data gathered on accidents, traffic citations, roadway characteristics, emergency medical services, and a host of other topics will be used in decision making whether or not the quality of the data is adequate. Although this statement certainly provides ample reason to improve data quality, a further analysis is required to obtain a full appreciation for the critical role played by data and the information derived therefrom in the decision-making process. Decision making in this sense refers to that process by which resources are allocated. The issues of traffic safety budget allocation are emphasized here, but the principles apply equally to any process in which scarce resources must be allocated to address a virtually infinite number of potential problems.

The statement that data will be used regardless of its quality is based purely on empirical observation of past practice. In the absence of other evidence on which to base a decision, the available data, albeit of the poorest quality, are generally given credence. Many pilot data collection projects suddenly turn into the real thing, and the data are used to prove hypotheses without any experimental design. Is this practice necessarily bad? Is it not possible that relatively good decisions have been made even though the data were not of the highest quality and were not gathered for the purposes to which they were later applied? These questions are addressed in the following paragraphs.

The only value of a vehicle crash is the information that it provides to decision makers to prevent a similar occurrence in the future. A crash costs anywhere from a few hundred to hundreds of thousands of dollars, depending on the severity. However, if it can be used to prevent other crashes, then it has intrinsic worth. However, this value may be completely lost if the quality of the data collected is inadequate. In fact, if the data are misleading, their use could be counterproductive. (This situation is not believed to be generally true, but practitioners should be aware of the possibility.) Finally, even good data that could be beneficial in guiding decisions can be neutralized by the questioning of its credibility. A series of premises are presented in the following section to set the stage for the proposed solution.

Department of Computer Science and Engineering, Auburn University, 107 Dunstan Hall, Auburn, Ala. 36849-5347.

PREMISES

Perfect Accuracy Is Impossible

The problems of data quality must be viewed in the proper perspective before attempting to formulate a conclusion. The major aspect of data quality is accuracy, defined in the context of traffic safety to include both completeness and validity of the data. Most traffic accident data elements are nominally coded data, that is, multiple choice questions. Invalid responses result from arbitrarily marking an incorrect response, misinterpretation of the question, data entry error, and other shortcomings of the data collection instrument itself. Incompleteness occurs when a data element is omitted, for whatever reason, and does not get recorded on the accident record. A second form of incompleteness occurs when a data element necessary for decision making does not exist in the data base. Another example of inaccuracy involves recording a milepost that varies from the actual accident location by a significant amount.

In a typical state data base of more than 100,000 traffic accident records per year, some inaccuracies are bound to creep in. The questions that should be asked are, What are the ramifications of these shortcomings? Does this inaccuracy alter policy to any degree? If not, there is no reason to worry about such errors. But if the deficiencies are altering policy—or, worse yet, if they are being used to discard the information value of the data altogether—then aggressive action is warranted to rectify these problems.

Perfect accuracy is unattainable, and those who insist on it are being unreasonable. No decision in life is based on total certainty. Why should the data on which safety decision making is based be any different? For example, election predictions are made with great accuracy on the basis of a small proportion of the results. If a good forecast can be made with less than 10 percent of the vote, should decision makers hesitate to use an accident records data base that is known to be at least 90 percent complete in most data elements? Of course, it is important that missing cases be randomly distributed throughout the population. The data base can then be viewed as a large random sample (proportionately speaking). If one geographic area or type of accident has more than its share of missing cases, adjustments can be made. However, the credibility of the remaining data does not automatically need to be questioned.

Another good example of sampling theory is in industrial quality control, which is a well-developed and sophisticated discipline. Products that require destructive testing for validation obviously cannot be subjected to 100 percent sampling. Sound inferences about the quality of a production run can be made with just a small percentage of the lot as a sample. This theory can certainly be applied in those states that monitor accidents by type as required by NHTSA for problem identification. Even though all accidents are not reported, sufficiently large numbers are available to ensure that the sample is relatively the same from year to year. Thus, when a particular type of accident (e.g., child pedestrian accidents in urban areas) is suddenly significantly higher in frequency than it was over the past few years, this situation can alert decision makers to take action.

Improvement in data quality will have little effect on the types of decisions for which current sampling is adequate. For these types of application (and possibly these only), the current quality is sufficient, and the data should be used and trusted. However, accident data bases are being pushed to produce more information of various types, which leads to the critical question, What quality is required? The answer depends on the application. The following table presents several applications with an assessment of the degree of accuracy required and available for each. These assessments are based on an estimate of the average state; any given state will probably vary from this norm. Clearly, a range of accuracy is required, and all data do not have the same quality requirement. With this concept in mind, it is important that the next premise be understood before proceeding.

Application	*Accuracy Assessment*
Problem identification	Currently adequate, Sufficient for Most Types
Hazardous location ID	Needs improvement but data are certainly usable
Location investigation	More details and greater accuracy would improve policy
Evaluation	Inadequate level of detail for many countermeasures

Too Much Data or Quality Can Be Counterproductive

Before the advocates of data quality become alarmed at this premise, they should recognize that too much of anything is, by definition, counterproductive. Thus, the argument reduces to determining whether or not it is possible to have too much data or too much data quality. To define terms, "too much data" refers to the collection of too many data elements, which can result in data collection forms that are too large or in data bases with too many records or too many types of records. These situations often occur in the name of completeness or because of the possibility that a certain application may occur in the future. The definition of quality, on the other hand, includes both accuracy and completeness. Finally, it is not being argued that there is currently too much data or that the quality is currently too high. The premise states only that there is a possibility that these two conditions can exist.

As stated, the purpose of collecting data is to guide the decision-making process in the allocation of limited resources. This objective leads to a paradox for the data base designer because data base design, data systems design, data collection, entry, maintenance, and ultimately processing to produce information all consume resources that could go into direct countermeasures. Could data (actually metadata, or data on data) be gathered to determine how many resources are to be allocated to the data subsystem? If so, should not an attempt be made to gather meta-metadata, or multimetadata? That this process of data collection could be extended infinitely proves the premise about data quantity to be true. The creation of even the first level of metadata is so expensive that most decisions about the content and structure of data bases are made on the basis of experience rather than hard data.

The data quality issue is more difficult to prove because there would seem to be an upper limit on data quality, especially for nominally coded variables, which make up most of the accident records data base. How much more difficult is it for an officer in the field to check the right code as opposed to the wrong code? How much more does it cost to measure the milepost to the nearest 0.01 mi as opposed to the nearest 0.1 mi? But it must be admitted that, no matter how hard one strives for perfect quality, it is always possible to do better. For example, the milepost reference system could be replaced with an alternative that would locate accidents much more accurately and reliably. Similarly, modifications could be made to most of the nominal-level variables to improve data quality. Are these improvements justified? In most cases, even minor enhancements cannot be made expediently because they will disrupt the system. Thus, compromises in quality are essential, which proves that there is an optimal level of quality and that it is somewhere on this side of perfection.

The practical ramifications of the premise are much more important than its proof. Too much data and an overemphasis on quality of certain unimportant data elements have resulted in a neglect of other data elements that are essential to decision making. When a data element is proposed, it should be recognized that the element must be properly defined, collected, entered, maintained, and processed if it is to be useful. A breakdown anywhere in this information-production chain can yield a data element that is counterproductive. Thus, an economic balance should be maintained, not just between data collection and direct countermeasures but also between competing data elements within a given data collection instrument.

The recognition that these two conditions (i.e., too much data and too much data quality) are attainable is important so that the objectives of the data-base design project can be put into perspective. Designers of data collection forms must recognize that they are responsible for the following: (a) determining the optimal number and type of data elements to maximize collector productivity, (b) designing these optimal data elements to produce the information required for effective decision making, and (c) providing guidance to the data collectors on the degree of accuracy required.

Although the proof of this premise would seem to oppose data quality advocates, such is not the intent. Rather, it is only by recognizing that an optimal level of quality exists that designers can deal with the necessary issues to achieve this level. The following premise addresses the current state of accident data as opposed to potential excesses.

Current Accuracy of Data Is Generally Not Sufficient

The qualifier "generally" must be emphasized before giving evidence of the validity of this premise. It is believed that a general negative attitude toward state and federal accident data currently exists, which is unjustified. (There is not a system in existence that cannot be criticized because it is not capable of doing something for which it was never designed. Current accident data systems are no exception. Rather than concentrating on the benefits that the data can produce, the focus is shifted to the vulnerable areas—things the systems cannot accomplish mainly because they were never designed for that purpose.)

Typical state accident records systems are now capable of problem identification, high accident location identification, and certain specific countermeasure evaluations. Increases in quality can bring about tremendous gains in these areas. However, as the data systems continue to be modified to produce greater data quality, these processing functions must not be put on hold pending better data that may never materialize.

Given this qualifier, the evidence supporting this section's premise is overwhelming. Is there a researcher who has not accessed the accident records data base to answer the simplest of questions and found that the needed variable was totally missing, not captured, not entered, or so misunderstood by the data collectors that its values were unusable? The unreasonable number of missing values and the gravitation to certain codes (e.g., inattention) are thorns in the side of those who are trying to turn data into information. Inconsistencies in the data abound in those variables, which are redundant in their potential information content.

These statements are based on first-hand experience with at least 10 different state records systems. Because there were no exceptions, these conclusions are not believed to be hasty generalizations. So, although there are some good data elements and a sufficient level of quality to generate some valid information for decision making, it will be a long time before the information benefits that optimal-quality state-level data are capable of rendering can be achieved.

If the consequences of inadequate data quality were limited to these problems, researchers could plot gleefully ahead toward greater quality. However, the problem tends to feed on the good data in an accelerating vicious cycle. Inadequate data quality in some data elements has led to limited application of all data. Legitimate applications are challenged due to known deficiencies. Most critical is the demoralizing effect on those charged with the use of the data. Why expend the effort to produce useful information from data known to be of inadequate quality? This question has too often led to deferring innovations until the data quality is improved, which is the key issue addressed here. Its counterproductive nature is discussed after the conclusions of the premises stated so far are presented.

CONCLUSIONS ABOUT PREMISES

Current Methods of Attaining Quality Are Not Sufficient

The following two approaches are used to improve the general quality of data obtained from state accident records systems: (a) appeals to altruism and (b) mandates. Because the quality of the data is inadequate, neither of these devices has been successful in motivating sufficient action. The following discussion concentrates on the data collectors as one source of the problem. However, other links along the data-to-information chain can be equally responsible for poor data quality. For example, poor forms design, inadequate edits, and flaws in data entry can be just as devastating to data quality as a neglect on the part of the data collector, who in most cases is a police officer.

Appeals to altruism are insufficient because competing activities may seem even more altruistic. The police officer will sacrifice a concern for detail (e.g., who was wearing what type of restraint by seating position) to remove the injured safely from the scene or to prevent a secondary accident by restoring traffic flow. In fact, the entire area of data collection is generally inconsistent with the paramilitary perspective that has been developed through years of training and experience. The police officer is trained to act, to take control of the situation, and to provide remedial and preventive effects by this action. Generally, police organizations are viewed by the public as being the custodians of expertise on traffic safety by virtue of their being on the scene. The idea that this experience can be coded into a computer and transferred to other organizations for information generation and dispensation represents a real loss of power in this regard.

Paramilitary organizations usually respond quite well toward mandates. To a large extent, the adequate data quality available in certain data elements is attributable to this factor. However, as the quality requirements increase, the officers begin to question why they need to provide this much detail and whether it is really necessary. Ineffectual data elements of unnecessary complexity can then become extremely counterproductive. Mandates are of little value here because they are impossible to enforce. In fact, they may bring a backlash of resistance if pushed too far.

Although police organizations have been used as the primary example, they are not intended to be a scapegoat. Data quality suffers in other parts of the system as well, and for the same reason, that is, current motivational methods are not working.

A New Approach to Quality Improvement Is Required

The only reason this conclusion might not follow from the previous one is the supposition that more of the same (i.e., more appeals to altruism or more mandates) might improve data quality. That these techniques have not produced the desired level of quality over the past 20 years is adequate evidence that a new approach is required. Instead of arguing this case further, an alternative called distributed responsibility is suggested. In this approach the data collectors (as well as the other weak links along the chain) assume respon-

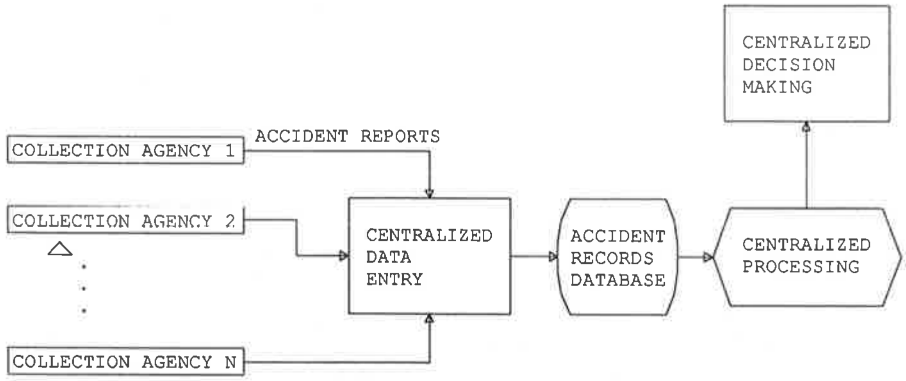

FIGURE 1 Traditional organization for centralized responsibility.

sibility for greater data quality. They do so because they find that it is in their personal or their organization's functional self-interest. Thus, the question of why so much detail is required is answered not by some mandating administrator but by the system itself.

A distributed responsibility system is data collector centered in that compellingly useful information is fed back to the organization that generated it. Although the examples presented focus on the law enforcement community, the applications are widespread. Because accident and citation information emanates from law enforcement, that community would be the target of these distributed responsibility benefits. However, in this era of integration, the same principles apply to medical data from hospitals, emergency medical data from emergency medical technicians, and even roadway characteristics information from engineers. The loop must be closed so that those responsible for data quality either get the full benefits of their labors or suffer the consequences of their deficiencies.

Figure 1 shows the traditional centralized responsibility model. In this paradigm each of the data collection agencies (e.g., state and local police) submit the data to a centralized data entry organization, after which the data go into the data base and are processed. The first attempt at distributed benefits is shown in Figure 2, where static hard-copy reports are sent back to the data collection agencies. (These agencies are collapsed into a single box to simplify the diagram.) Clearly, an alternative term for distributed responsibility is distributed benefits, because it is the benefits that will motivate the entire organization to assume proper responsibility. It is essential that the data collectors see the output and recognize its value. Traditional outputs—standardized hard-copy reports—have failed to provide this motivation because they have failed to obtain the participation of the data collection organization in their generation and use.

These two aspects of the process are inseparable: generation and use. Standardized hard-copy reports seldom answer the detailed questions that arise dynamically in the field. Rather, they tend to represent the only output of the process to the data collectors. Thus, if the specific required information is not in these static reports, the conclusion is that the entire process (data collection through output generation) is worthless, which further undermines the data quality for currently useful applications.

Distributed benefits must be technology based. These capabilities were not possible in the era of hard-copy outputs. Now, however, with the ubiquitous nature of personal computers, there is no reason that software to generate information useful to the local organizations cannot be given to them (costs are discussed in a subsequent section). This capability, coupled with the advent of new and advanced input technologies, can lead to an order of magnitude increase in

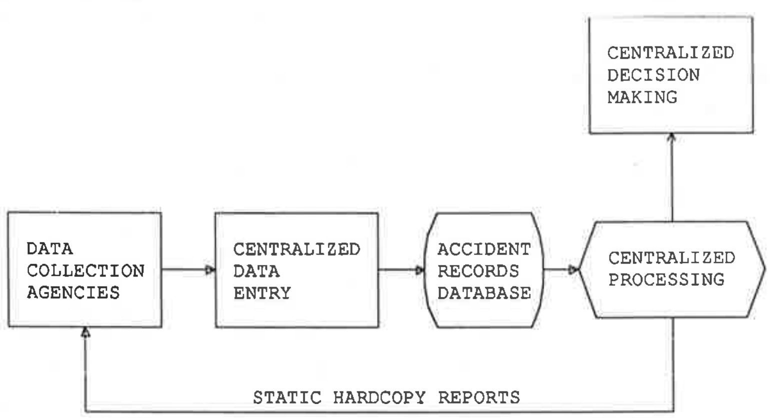

FIGURE 2 Initial attempts to provide distributed benefits.

the quality of data available throughout all organizations in the traffic safety community.

EXAMPLE IMPLEMENTATIONS OF DISTRIBUTED BENEFIT SYSTEMS

A number of distributed benefit systems exist, none of which has reached its full potential of user orientation and application. The purpose of this section is not to promote any given system or approach; rather, it is to promote the concept of distributed benefit systems in general. All of the systems mentioned, and the many others believed to be currently under development, should strive to provide local decision makers with information, not data. This information needs to be in a form that can be understood and used directly by the decision maker, recognizing that statisticians and computer technicians might not be available to local organizations.

A list of distributed systems is given in the Appendix. The list is not purported to be complete, and other systems will continue to be announced. However, these systems do represent a fairly good cross section of the current state-of-the-art. Table 1 presents a comparison of the various aspects of these systems as of March 1990, when several of them were presented in a NHTSA-sponsored conference (*1*). (Because these systems are in continual development, the representations are subject to change. The contacts given in the Appendix can be consulted for current information.)

Table 1 indicates that the current systems cover a wide range of data sources and entry responsibility. Most of the local systems require their own data entry. The demand for such systems indicates that a large number of them are already developed at local levels but have not been publicized. These systems will provide benefits by increasing data quality, especially for those data elements maintained at the local level. They will also promote awareness of the value and use of data that are sent to the state from local areas. A further benefit is the availability of the most current data to the local officials.

Figure 3 shows the locally autonomous systems. There are two independent data entry functions, which may be inefficient. Data entered at the local level are usually only a small subset of the total record. However, because the state and local data entry are redundant, there may be a problem of consistency between the two. There is also a question of whether the local concern for the data sent to the state is increased or decreased, because they may reason that their data needs are now satisfied. These problems are not insurmountable; the local data entry systems are just a stepping stone to that point in time when all data will be entered locally and uploaded to the state, thus eliminating redundancy and greatly increasing the quality of all data.

As indicated in the second entry of Table 1, a few states are providing a downloading service to furnish information capabilities to the local organizations. This model, as shown in Figure 4, is fully capable of returning a complete data access capability to the locality without redundant data entry. The model has the advantage of efficiency and consistency, but it is resisted by some jurisdictions because it suffers in data timeliness. However, because it gets a turnkey system into the hands of local officials in the quickest possible time with the largest possible data base, the timeliness shortcoming might be overlooked in the near term. Once the local data entry capability is established, the responsibility can be transferred locally (again, preferably in the direction of ultimate automatic uploading of locally entered records).

The third level of sophistication is with those states that already enter data locally and upload to the state level, as

TABLE 1 CLASSIFICATION OF DISTRIBUTED SYSTEMS

Classification	Systems
Autonomous: no interaction with state data base	MTRS SCARS KARS TARS
State-Dependent: Data downloaded from state	CARE TARS
Local-Entry–Dependent: Data uploaded to state	LANSER TRASER
Totally Local–Dependent	None currently known

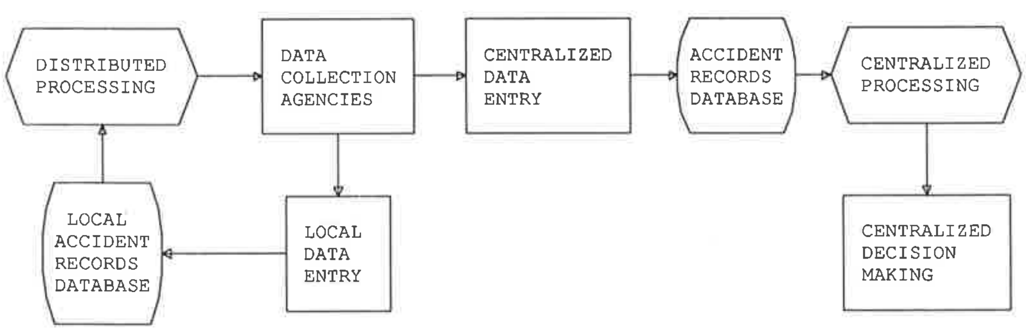

FIGURE 3 Locally developed and implemented microsystems.

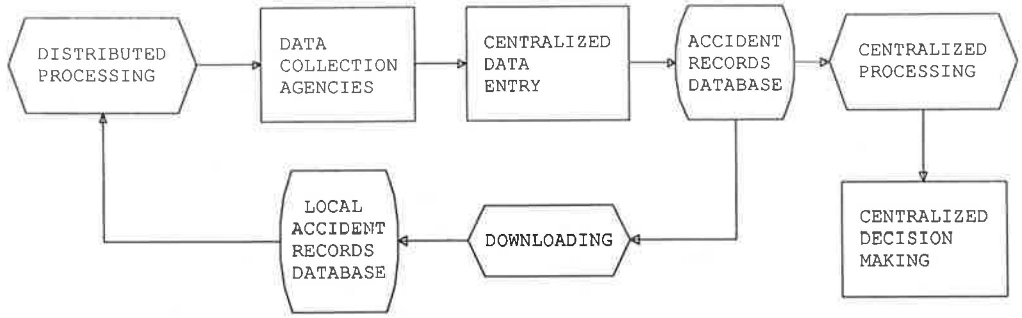

FIGURE 4 Distributed responsibility/benefits model.

shown in Figure 5. Here, networking capabilities are being exploited on a pilot basis, and the concept is proving to be feasible. Although no examples are given of a totally locally dependent system, a representative at the San Diego meeting (*1*) indicated that this was the case in Ohio. In this extreme, the local organizations own the data and develop their own processing capabilities. This method seemed satisfactory, but there clearly, is an optimal point of distribution of responsibility after which inconsistency between local jurisdictions can become counterproductive. All of these new distributed systems are still attempting to find that optimal point.

CURRENT DISTRIBUTED SYSTEM CAPABILITIES

The basic objective of the distributed microsystems is essentially the same: to provide local access to information for local decision makers who have little, if any, formal computer or statistical training. A further objective is to minimize the implementation expense. Therefore, these systems have been designed to operate on hardware that is largely available (mainly IBM PC/AT compatible). They all have the capability to create subsets of the accident records so that such types as alcohol, bicycle, and pedestrian can be examined. This capability ranges among the various systems in both sophistication and user friendliness.

Current network technology makes it feasible to enter all data from the local level. The most basic requirements for electronic communication between a remote site and the central data base include a modem and a telephone connection. The modem is a simple device that translates digital signals from a computer to and from analog signals carried by a telephone cable. With the addition of a dedicated phone line and modem, two remote computers can communicate easily.

Even when all police officers have personal lap-top computers in their squad cars, the process is not complete until the central data base is updated. Initially, this task will be accomplished by storing the data gathered by the officer onto a diskette, placing the diskette in the local office computer, dialing the central computer, and uploading the information. Information can be disseminated to local officials in the reverse fashion. Data collected and processed in the central location can be sent electronically over the telephone system to the remote sites.

A complete networking of all computers in a state's traffic control offices is an achievable goal. A centralized data base will then be available to all sites, and data-base updates will be made with no replication of data entry and no conflicting data. For a region's traffic data, a viable distributed system includes a local area network (LAN) with a central data base acting as a server for the remote clients seeking information. Data base updates will be available immediately to all servers on the network, and all servers will be able to modify the data base directly with their local input. Figure 6 shows such a distributed system. Although this model is currently not being implemented, the technology is currently available, and it is expected to be the direction that distributed state systems will take over the next decade.

The basic output of current systems is the frequency distribution for a given variable or set of variables. Some have standardized output report generators, which limit the users to predefined outputs. Although these outputs have little more than hardcopy flexibility, that they can be obtained for virtually any subset is a definite advantage over the model given in Figure 2. However, most have the capability, at least for

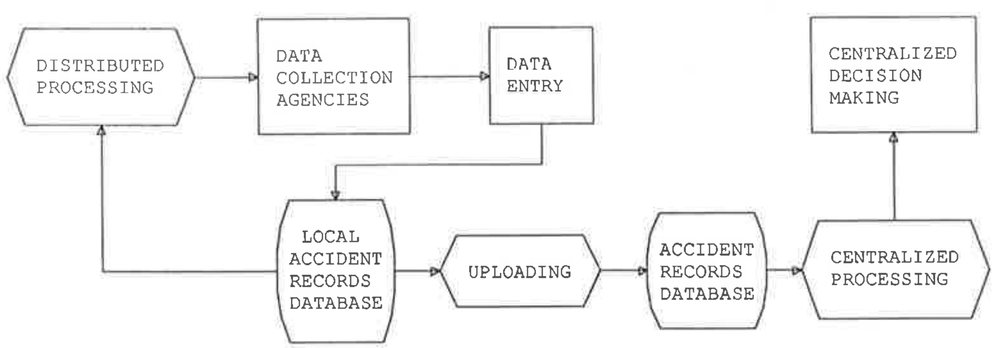

FIGURE 5 Distributed system under current technology.

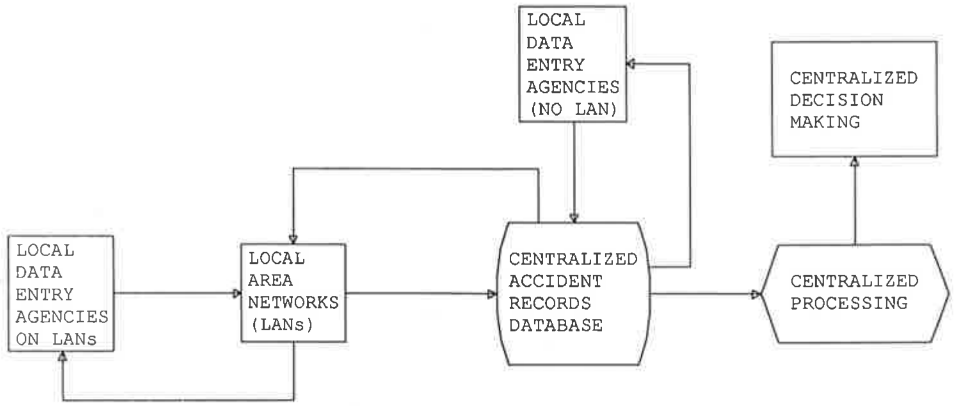

FIGURE 6 Networked distributed system.

the more advanced user, to bypass the standard reports and generate virtually any summary of information from the data base. Some are based on dBASE or can generate ASCII files so that their capabilities are extensible. The generation of any cross tabulations and the automatic identification of overrepresentations are two capabilities that are of major value.

High-accident locations can be generated easily on most of these systems using either user-specified criteria or the specification of an exact location. Additional information on a location can also be obtained by frequencies, cross tabulations, or, in some cases, standardized reports. For those wishing to retrieve the hard copy of the accident report, the accident numbers are readily available.

The use of graphical outputs is just beginning to find its way into local systems. Bar charts from frequency distributions and color within cross tabulations enable the users to visualize overrepresentations in a much more understandable way than viewing a list of numbers. Some of the systems can automatically generate collision diagrams, which seems to be feasible and should have widespread application within the next year or so.

As mentioned, the cost of a distributed system is minimal when the local departments are already equipped with computers, even if these are only IBM PC–compatible microcomputers that are being used primarily for word processing. Although some small police departments might not have this technology yet, it is expected that they will soon. With such equipment, the only hardware requirement is a centralized PC file server to store and protect the centralized data base, as well as modems at each remote PC. The server must contain a large (minimum 1.2-gigabyte) hard disk, and it must be accompanied by a backup system (possibly a tape drive costing approximately $6,500). The modems are required for data transmission over telephone lines, and the typical 9,600 bits/sec modem would cost about $400. If several PCs are employed in a larger department, it would be preferable to have them connected by a LAN. This would require the purchase of Ethernet cards for each PC (approximately $150 each) and server software, which can range in price up to $6,000.

Table 2 presents an approximate price list based on the establishment of a LAN consisting of five PCs at one locality. The costs in this table assume that the department effectively starts from scratch. Most local departments would not have this system dedicated to their accident records. They might be using it for departmental administration, citation processing, word processing, and a variety of other applications. Thus, only a small fraction of this cost would need to be justified by the accident records system. In many cases, this equipment is already installed for these purposes, and only a small marginal cost is required to install the accident-processing components. Such is also the case with a small department that already has a single PC with time and space

TABLE 2 POTENTIAL SYSTEM COSTS ($)

ITEM	UNIT PRICE	QUANTITY	TOTAL
386 PC & network card	2500	5	12,500
386 server	6500	1	6,500
Laser printer	3500	1	3,500
LAN software pkg.	6000	1	6,000
Modem	400	5	2,000
Coax cable (1500 ft)	500	1	500
Misc. Network Hardware	500	1	500
Labor			7,500
Training			3,500
TOTAL			42,500

availability and therefore only require a modem and a communications card.

CONCLUSIONS

The major conclusion obtained from considering these premises and their supporting facts is that innovations in technology applied to the entire data-collection–to–information process will create data quality, provided the human element is not neglected. The greatest technology advances will not render the data more usable if the data collectors and other support personnel are not properly motivated. The success of distributed responsibility depends on distributed benefits. It is essential that the data collectors and their respective local organizations be targeted for software designed to be used as an integral part of their jobs, which will give them a vested interest in seeing that the data are accurate and complete. In the meantime, prototype data should be used to continue innovation. Data inaccuracy should never be used as an excuse to delay innovation, for it is only by having the capability to generate benefits that the value of quality data can be appreciated.

ACKNOWLEDGMENTS

The research and findings presented in this paper were partly sponsored by the Alabama Highway Department, the Alabama ADECA Traffic Safety Section, and NHTSA.

APPENDIX
Contact List of Micro-Based Accident Systems

City Accidents Rapid Evaluation (CARE)
David B. Brown
Department of Computer Science and Engineering
Auburn University
107 Dunstan Hall
Auburn, Ala. 36849-5347
(205) 844-6314

Kansas Accident Records System (KARS)
University of Kansas Transportation Center
PC-Transmission
2011 Learned Hall
Lawrence, Kans. 66045

Local Area Network Safety Evaluation and Reporting System (LANSER) and Traffic Services Microcomputer System (TRASER)
Marlin Crouse and Barbara Delucia
Texas Transportation Institute
Texas A&M University System
College Station, Tex. 77843-3135

Small Computer Accident Records System (SCARS)
Ken Courage
McTrans Center
University of Florida
512 Weil Hall
Gainesville, Fla. 32611

Traffic Accident Reporting System (TARS)
Steve Lau
Lau Engineering, Inc.
17220 Newhope Street, Suite 204
Fountain Valley, Calif. 92708-8771
(714) 546-2046

Virginia Micro Traffic Records System (MTRS)
Robert Breitenbach and Jill F. Davis
Transportation Safety Training Center
Department of Justice and Risk Administration
816 West Franklin Street, Box 2017
Richmond, Va. 23284-2017
(804) 367-6235

REFERENCE

1. *Proc., Conference on the Collection and Analysis of State Highway Safety Data*, San Diego, Calif., NHTSA, U.S. Department of Transportation, June 1990.

The opinions expressed in this paper are those of the authors and do not necessarily reflect positions held by the sponsors.

Publication of this paper sponsored by Committee on Traffic Records and Accident Analysis.