U.S. Department
of Transportation

**Research and
Innovative Technology
Administration**

Bureau of
Transportation Statistics

DOT Report for Implementing OMB's Information Dissemination Quality Guidelines

THE SECRETARY OF TRANSPORTATION
WASHINGTON, D.C. 20590

SECRETARY'S POLICY STATEMENT ON INFORMATION QUALITY

August 2002

Information is a critical Department of Transportation (DOT) resource, second only to human resources. It is vital not only to DOT's daily operations; it is an essential element in fulfilling our mission to ensure the safe, effective, and secure operation of the entire transportation system. Further, in the course of our work, we generate a wide variety of information and information products for public use. For example, we issue statistical reports, studies of important safety and other transportation issues, analyses of the costs and benefits of regulations and policies, scientific reports, environmental assessments, and many other documents. As public servants, we are obligated to ensure that all DOT information products consistently meet or exceed high standards of quality.

Section 515 of the Treasury and General Government Appropriations Act for FY 2001 directed the Office of Management and Budget (OMB) to issue government-wide guidelines to Federal agencies for ensuring and maximizing the quality (objectivity, utility and integrity) of information disseminated by Federal agencies.

Consistent with OMB's guidelines, the Department has drafted agency-wide guidelines to establish and apply high standards of quality (objectivity, utility, and integrity) to government information prior to public dissemination. The guidelines *(which can be found on DOTNet, under "On-line Policies")* also guarantee affected members of the public the opportunity to request correction of perceived misinformation. Each operating administration is encouraged to further refine guidelines to reflect the nature and sensitivity of its specific information products.

I am urging all employees to make a strong commitment to apply high standards of quality with respect to all of our information products and services. Only by making information quality a regular part of our daily work can we fulfill this important obligation to the public. I am also calling upon senior managers, program officials and supervisors to provide the leadership needed to implement these guidelines. We must make information quality a high priority in order to fulfill our responsibilities to the American taxpayers.

Norman Y. Mineta

**The Department of Transportation's
Information Dissemination Quality Guidelines**

WHAT IS THE PURPOSE OF THIS POSTING? Consistent with The Office of Management and Budget's (OMB) Guidelines (for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies) implementing Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (P.L. 106-554), the Department is issuing guidelines explaining how the Department will ensure the quality of disseminated information. This document also explains how affected persons may seek and obtain corrections of information that does not comply with these guidelines.

WHEN ARE THESE GUIDELINES EFFECTIVE? These Guidelines are effective October 1, 2002.

WHO SHOULD BE CONTACTED FOR FURTHER INFORMATION ABOUT THESE GUIDELINES? Steven B. Lott, Office of the Chief Information Officer, U. S. Department of Transportation, 202-366-1314 (not a toll-free call) or by e-mail at Steven.Lott@ost.dot.gov. For inquiries on the Department's administrative mechanisms for persons to seek correction of information, please contact Robert Ashby, Office of the General Counsel, U. S. Department of Transportation; 202-366-9306 (not a toll-free call) or by e-mail at bob.ashby@ost.dot.gov. For inquiries on the guidelines concerning statistical disseminated information, contact Dr. Patrick Flanagan, Bureau of Transportation Statistics, U.S. Department of Transportation; 202-366-4168 (not a toll-free call) or by e-mail at Pat.Flanagan@bts.dot.gov.

COMMENTS AND DOT RESPONSES:

In response to its May 1, 2002, posting of the draft guidelines, the Department received eight substantive comments from members of the public. The commenters were the U.S. Chamber of Commerce (COC), the Center for Regulatory Effectiveness (CRE), Citizens for Sensible Safeguards (CSS), the American Bar Association Section on Administrative Law and Regulatory Practice (ABA), Senator Richard Durbin, the National Association of Manufacturers (NAM), the American Trucking Association (ATA), and Senator Joseph L. Lieberman. These commenters discussed a variety of issues. In the discussion of these comments, as well as in the DOT guidelines themselves, when we use the term "guidelines" we mean both the OMB guidelines and the DOT guidelines.

<u>Scope of the Guidelines</u>

COC, CRE, and NAM asserted that the guidelines should be viewed as legally binding rules. CSS asserted the contrary. It is clear from Section 515 and from the OMB guidelines that the purpose of the agency guidelines is to set forth policies, procedures, and recommendations for ensuring information quality. (It is our understanding that an

earlier, never-enacted draft of the amendment that became Section 515 called for "regulations," a term replaced by "guidelines" in the final version). Congress is very experienced in creating mandates directing agencies to create legally binding regulations. Such provisions are a regular part of statutes that DOT and other agencies are required to implement. Where Congress chooses a different approach, as it did in Section 515, DOT would be acting contrary to statutory direction if it attempted to convert the intended flexible guidelines into hard and fast, legally binding rules. We have retained the "guidelines not rules" structure and language of the proposed guidelines.

COC suggested that DOT operating administrations should have a notice and comment process to modify or augment these DOT-wide guidelines. It is not clear that DOT operating administrations will ever issue their own equivalents of this document. What is more likely is that the operating administrations will determine internally how to apply DOT's guidelines to their own programs and information products. If it is useful for an operating administration to obtain public comment, or if an operating administration wished to depart significantly from the approach the DOT guidelines take, the operating administration could issue a draft for comment. Whether that course will be appropriate will be determined as the process of implementing these guidelines evolves.

Exceptions to Coverage

CRE asserted, per a lengthy legal memorandum provided to all the participating agencies, that it was contrary to statute for OMB or the agencies to make any exceptions to the definition of "dissemination" in the guidelines. The Department is not persuaded by CRE's analysis.

CRE's argument begins with the premise that Section 515 is an amendment to the Paperwork Reduction Act's (PRA) information dissemination requirements. This premise is incorrect. Section 515 is a free-standing statutory provision, which was assigned as a classification matter for purposes of codification as a "note" to 44 U.S.C. §3516, a portion of the PRA. A "note" does not amend the section or overall statute to which it is assigned. The assignment of the note to a particular section simply denotes that the codifiers decided upon a convenient place to insert the provision. The only direct connection between Section 515 and the PRA is that the Section 515 directs OMB to apply its guidelines to agencies that are subject to the PRA.

CRE is forced to take an indirect route to its conclusion because the text of Section 515 imposes no limits on OMB's discretion concerning exceptions to the definition of dissemination. There is, of course, little if any legislative history available for Section 515. In these circumstances, neither OMB nor other agencies confront a direct statutory limitation on their normal discretion to interpret statutes in a reasonable manner. It is reasonable to assume that Congress knows how to prohibit agencies from exercising regulatory discretion; Congress shows no sign of doing so in Section 515.

The thrust of CRE's position is that because other applications of the term "dissemination" (e.g., in OMB Circular A-130, the legislative history of portions of the

1995 amendments to the PRA) are broad, and do not have exceptions similar to those in the OMB guidelines, Congress intended to prohibit exceptions to "dissemination" in the context of Section 515. While CRE's argument is an interesting exercise in bootstrapping, it does not meet applicable legal tests for constraints on agency discretion.

Chevron U.S.A. Inc. v. Natural Resources Defense Council, Inc., 467 U.S. 837, 843, 81 L. Ed. 2d 694, 104 S. Ct. 2778 (1984), and its progeny provide the applicable standard of review in matters of agency interpretation of its discretion under a statute (it is important to note that the *Chevron* analysis is not limited to rulemaking matters). Under the *Chevron* analysis, judicial review of an agency's interpretation of a statute under its administration is limited to a two-step inquiry. At the first step, courts inquire into whether Congress has directly spoken to the precise question at issue. If a court comes to the unmistakable conclusion that Congress had an intention on the precise question at issue, our inquiry ends there; the court naturally must give effect to the unambiguously expressed intent of Congress. However, if the statute before the court is silent or ambiguous with respect to the specific issue, the court should proceed to the second step. At this stage, the court defers to the agency's interpretation of the statute if it is reasonable and consistent with the statute's purpose; the court is not free to impose its own construction on the statute, as would be necessary in the absence of an administrative interpretation.

As noted above, Section 515 contains no explicit, or even implied, prohibition on OMB or other agencies with respect to their discretion to make reasonable exceptions to the definition of dissemination in information quality guidelines. References to the 1995 PRA amendments and their legislative history are, at best, a very weak substitute for actual evidence that Congress intended to prohibit OMB and the agencies from acting reasonably with respect to use of "dissemination" years later when it enacted Section 515. Consequently, under the first part of the *Chevron* test, Section 515 is silent or ambiguous with respect to this issue. Under the second part of the test, it is clear that the exceptions that OMB and DOT have chosen to express are reasonable and advance the statute's purpose of improving information quality. Insisting on a "no exceptions" approach, while imposing unnecessary additional administrative burdens on agencies, does little to actually improve information quality in the areas of greatest concern to the public.

COC suggested that, with respect to correspondence sent to individuals, information that DOT has reason to believe would be distributed further by the recipient should be covered. For example, if DOT sends a letter to the head of an interest group, and we anticipate that the group will send copies to its members, we should apply the guidelines, notwithstanding the correspondence exception. The Department has not adopted this comment. The decision to distribute a letter is the recipient's not DOT's. The recipient's decision to do so does not trigger further obligations on DOT's part. Only if DOT sponsored or requested the further distribution (e.g., sends a letter to an interest group that requests the group to distribute the letter widely) would we agree that the correspondence exception would not apply.

The ABA suggested that we significantly limit or delete the "public filings" exception, such that material in a DOT docket would be covered by the guidelines. It appears that

the ABA may have misunderstood the DOT's proposal in this regard. What DOT proposed, and what we have adopted in these final guidelines, is a provision that does apply the guidelines to docketed material, if and when the Department uses and disseminates the material. However, the mere fact that information is sitting in the docket, untouched by Departmental hands (aside from DMS personnel, perhaps), should not trigger applicability of the guidelines.

NAM disagrees with the proposal's handling of the exceptions for press releases and submissions to Congress, suggesting that the exceptions should apply only to information that has previously been disseminated subject to the guidelines. With respect to press releases and other ephemeral information products, DOT believes that it is difficult, and not very productive, to try to fit very short-lived, brief products of this sort into the guidelines' review and correction scheme. With respect to submissions to Congress, the Congressional legislative and oversight processes provide for a proven, effective means of correcting information that members of Congress believe to be flawed. A decision on a legislative provision can itself be a definitive judgment by Congress on the quality of information supporting or opposing the provision. While factual information presented to Congress (e.g., a study referred to a submission to a Congressional committee) that had not previously been subject to these guidelines would be subject to the administrative corrections process like any other information. Much material presented to Congress consists of agencies' policy views or information developed quickly in response to Congressional inquiries for which coverage under these guidelines is less apt.

CRE said that the definition of "affected person" should be broad, so as not to unnecessarily exclude persons who seek correction of information. NAM also suggested broadening of the term. The Department has not erected any elaborate "standing" requirement for seekers after correction. However, the statutory term "affected" person must be given some meaning; there must be some persons in the world who are "unaffected" if the term is not to be mere surplusage. Someone whose involvement with the information in question is merely that of a gadfly, bystander or interested reader, for example, could be an "unaffected" person.

In these guidelines, we use the term in the commonly understood way to mean a person who can show that he or she can be harmed by information that does not comply with the guidelines or can benefit from a correction of such information. We do not believe that it is useful to create an elaborate abstract definition of the term. Rather, in the guidance for the administrative corrections process, we ask requesters to explain how they are affected and provide discretion to DOT organizations to determine if the requester is an affected person who should receive a substantive response.

ABA said that if data is first disseminated before October 1, 2002, and retained on a Department's web site thereafter, it should be covered by the guidelines, evidently on the theory that it is continuously disseminated because it is still available to readers. In the Department's view, the ABA's position would give unreasonably retroactive effect to the guidelines with respect to documents that may be years old and were not subject to the pre-dissemination review provided in the guidelines. Maintenance of information on

4

DOT web sites or paper files does not, in itself, subject information disseminated before this date to the guidelines. However, the Department's policy is to treat as subject to the guidelines information that we maintain in a way that is readily available to the public and that continues to play a significant, active role in Departmental programs or in private sector decisions. NAM essentially agreed with the Department's position on handling of data first disseminated before October 1, 2002.

Rulemaking-related Issues

COC and CRE said that the guidelines should not exclude rulemakings. They do not. Since there was evidently some misunderstanding on this point, the final guidelines make this point even more explicit than the proposal. The substantive requirements of the guidelines (objectivity, integrity, etc.) apply to information in rulemakings just as they do to other information. What is different is the mechanism used to respond to requests for correction. As in the proposal, the final guidelines state that the existing mechanism of a response to public comment on a notice of proposed rulemaking (NPRM) in the preamble will be the primary way in which the Department responds to requests for correction about information in the NPRM. CSS said that information in rulemaking documents should be excluded from the guidelines. We do not agree that this would either be desirable or consistent with the statute and OMB guidelines.

Commenters did make three useful points about the use of rulemaking and analogous mechanisms for responses to requests for correction. First, as COC pointed out, if an NPRM came out before October 1, 2002, with the final rule published after this date, the public would not have had the opportunity to comment with respect to conformity of information to these guidelines. For this reason, we have changed the language of the guidelines to state that we would reject a request for correction only if someone had previously had the chance to comment on a document with respect to compliance with the guidelines. This is a transitional issue that we expect would not arise frequently after the first year or two of our implementation of the guidelines.

Second, COC and CRE said that a problem with responding to comments in a final rule is that final rules can come out a very long time after a proposal. We agree that, if it appears to the Department that a final rule will be delayed a long time, or the requester has made a good case that a quicker response is necessary, the Department could exercise its discretion to issue a response before the final rule is issued. We have added language to this effect.

Third, COC also asked how the reconsideration process applied to rulemakings. In the case of information in an NPRM, the response of the Department in the preamble to the final rule would be the initial response. Normally, a request for reconsideration of the initial decision on the request for correction would be handled either as a request for reconsideration of the rule or amendment of the rule under agency guidelines. However, in the case where the requester sought a correction that did not affect the rule – and the agency agreed that a correction would not affect the rule – the Department would follow through the same reconsideration process applicable in non-rulemaking situations.

In terms of deciding whether to accept a request for correction from someone who could have commented on an NPRM, ABA urged that the Department distinguish between commenters and (especially unsophisticated) non-commenters, giving greater consideration to the latter. Everyone has the chance to comment on a rulemaking. Those who are unaware of the rulemaking or choose not to participate do not have their voices heard, whether with respect to the contents of the NPRM generally or with respect to information quality issues. We do not know of any means by which we can distinguish the relative sophistication of commenters about the rulemaking process.

Third-Party Information

COC asked that the guidelines clarify that information submitted by contractors and grantees, as well as commenters, may be subject to the guidelines. We agree, and have added another example. COC and CRE also ask that third-party proprietary models seldom or never be used, and that strong robustness checks be employed when they are. DOT agencies seldom, if ever, use third-party proprietary models. Robustness checks in this situation are called for by the OMB guidelines, and we have added language to this effect.

ATA asked that the guidelines specifically apply to the Federal Motor Carrier Safety Administration (FMCSA) "SafeStat" system. This is a web-available ranking system for the safety performance of motor carriers. The inputs for the system include data from state as well as FMCSA inspectors. ATA is concerned that if some of the data from state sources is erroneous, there is no present mechanism by which FMCSA can correct it. On the other hand, the state-generated information, while used and disseminated by FMCSA, is not owned or in the hands of FMCSA, making correction problematic at best.

We agree with ATA that data quality of the SafeStat system is important and that the system, as a general matter, is covered by these guidelines. However, like other information systems in which the Department inputs and then makes use of data supplied by the states, FMCSA does not own, possess, or control the state-generated data. This makes application of the guidelines' administrative correction mechanism difficult.

Suppose, for example, that a motor carrier contacts FMCSA and says that the SafeStat web site contains inaccurate information about the carrier (e.g., the system says there was a crash involving one of the carrier's trucks and the carrier denies it). The information came from a local law enforcement agency to a state safety agency, which in turn uploaded the information to FMCSA. How would FMCSA staff in Washington, or agency contractors in Massachusetts, determine whether the report is accurate? If information appears inaccurate, through what process could FMCSA, which has no authority over the law enforcement agency, get the agency to change the report if FMCSA believed it to be incorrect? If FMCSA changed the report in the Federal database, then there would be an inconsistency between the Federal and state/local data, as well as a potential disruption in the necessary close working relationship among the Federal/state/local parties involved. FMCSA can work with state partners to improve the

6

transparency of data, include appropriate disclaimers in relevant databases, and correct some individual items where feasible. However, it would be misleading to portray these guidelines as a "magic bullet" that can effectively correct allegedly incorrect individual items in FMCSA or state databases on an across-the-board basis.

From a study FMCSA did of randomly selected motor carriers, it appears that the rate of errors in SafeStat data is small – on the order of ¼ of one percent. In addition, FMCSA already has in place some safeguards that fall under the general heading of pre-dissemination review, such as checking for discrepancies in incoming data and confirmation analysis of data prior to release. Nevertheless, FMCSA agrees that it has responsibility to facilitate the correction of the relatively infrequent errors in system data.

For this reason, FMCSA is currently developing additional data correction guidelines. The concept behind these guidelines is that FMCSA will provide more tools to the state to facilitate correction of motor carrier inspection and crash data. When motor carriers contact FMCSA with a concern about data accuracy, FMCSA will provide the information to the appropriate state Motor Carrier Safety Assistance Program (MCSAP) office, which would seek to resolve the matter and then notify the carrier of its action. Currently, states must re-upload the data to FMCSA's Motor Carrier Management Information System (MCMIS) in order to effect the change. However, FMCSA is enhancing MCMIS to allow state MCSAP offices direct access to make any needed changes in crash or inspection information. One action the MCSAP office could take with the enhanced system is to place a "hold" on disputed data so that it would not be included in the SafeStat system or other databases until the dispute is resolved. In the future, FMCSA will look at the possibility of automating this process further, perhaps through the use of an on-line correction request form that could be transmitted electronically to the appropriate state office. Before FMCSA issues final data correction guidelines, additional procedures will be developed and tested to ensure that data in error are corrected in a timely fashion.

The application of the guidelines to the SafeStat system is an example – the only one cited by commenters – of what may be a broader long-term issue. There may be a variety of programs in which the Department, or other Federal agencies, necessarily rely on or further disseminate information provided by state and local governments. Because of the action the Department takes with respect to the information, it appears that the information is subject to the guidelines. Yet application of the specific features of the guidelines may be very difficult. The Department, as it implements the guidelines over time, will work on ways of applying the principles and objectives of the guidelines in these situations, in ways that fit the individual circumstances of the programs involved.

Administrative Corrections Process

CRE and NAM asked that the deciding official for initial correction requests, as well as requests for reconsideration, be outside the office or organization that generated the information involved. Objectivity and a lack of personal investment in an information product are desirable qualities in a deciding official. So, however, is knowledge of the

subject matter. Admittedly, these two qualities can be difficult to find in the same individual. The guidance to DOT organizations is to do the best they can to find such a combination. However, mandating that the deciding official be outside the organization that produced the information could guarantee that the decision maker was unfamiliar with the issues involved. At the reconsideration stage for requests involving influential information, the final guidelines specify a three-person panel that typically would employ two persons from other DOT organizations.

Several comments addressed timing. CRE thought that there should be a maximum 45-day response period to requests for correction, while CSS thought there should be a 3-month statute of limitations on requests for correction after the dissemination of information. CSS also thought there should be a disclaimer that responses should be limited to information available at the time of the dissemination, so that there would not be a "moving target" situation.

We believe that, given the press of work at the Department, establishing a 45-day response period might be promising more than we could reasonably deliver. We have modified the 90-day time frame of the proposal to be 60 days, though we will strive to respond to requests as soon as possible in any case. We also believe a 3-month statute of limitations is unreasonably short, and ignores the problem, raised by other commenters, that the importance of a piece of information may not be readily apparent when it is first disseminated. Therefore, we have retained the proposal's suggestion that the Department can choose not to make corrections of information that is more than one year old, unless the data remains significant in ongoing DOT or private sector activities and decisions. In the final version of the guidelines, we have made this point a consideration in the Department's decision on a request rather than a "filter" that would screen requests out of the system.

We agree with CSS that the Department should not have to face a "moving target" situation, and we have added language to this effect. ABA suggested that we put a "flag" in a web site to note information that we have agreed to correct but have not yet corrected. While we have not included language to this effect in the guidelines, we will consider, as we implement the guidelines, whether this idea is practical and desirable.

Influential Information

COC, CRE, NAM and ABA objected to the use of the $100 million criterion in the proposal's discussion of influential information. CRE and NAM also believed that the discussion was too restrictive in its treatment of influential. We have clarified the discussion of the $100 million figure: if something has an impact of $100 million or more, the impact is likely intense, but having an economic impact of that amount is not essential to finding an intense impact. Depending on the circumstances and parties involved, a lesser economic impact could also contribute to a determination that information is influential.

We do not believe that the discussion of this concept is too restrictive. Indeed, it is an excellent example of the guidance orientation of this document, being a discussion of a thinking process designed to help DOT officials make sound decisions on the issue of influential information. In our view, consideration of the breadth as well as the intensity of an impact is essential to a fair decision on whether information is influential. In the absence of such consideration, virtually any information of interest to a particular party could be declared influential, which would be inconsistent with the intent of the guidelines.

Objectivity

With respect to peer review, CRE asked the Department to address the standards for rebutting the presumption of the adequacy of peer-reviewed information, while CSS suggested that we address the limitations of peer review. We have incorporated language in response to both suggestions.

COC, CRE and NAM commented that the guidelines should specifically adopt or adapt (preferably the former) Safe Drinking Water Act (SDWA) standards for reviewing influential information in the context of risk analysis. CSS suggested adapting rather than adopting.

While the Department may disseminate information products that evaluate safety risks in various ways, the Department appears not to conduct many formal "risk analyses" of the kind often promulgated by health and environmental agencies. The SDWA model is probably not a close fit for most DOT activities. Nevertheless, the general principles of the SDWA have wide application, and the Department will adopt or adapt these principles, as applicable, to its evaluations of risks. The final guidelines call on DOT organizations to apply the SDWA standards to their risk-related products.

Other Comments

CSS suggested that we add language placing compliance with the guidelines in the context of other agency objectives, such as carrying out agency missions, staying within budget and priority restraints, and providing useful information to the public. We think this is a useful suggestion, and we have added such language.

NAM said that the guidelines should identify who the appropriate data quality official is in a DOT organization. The Department does not believe it would be productive to list these operating administration contact persons by name in the guidelines, as their identity will no doubt change from time to time. The Department will consider having a current directory of these operating administration contacts on its web site, however. It should be pointed out that the role of this official is not to make determinations on requests for correction; it is simply to act as a facilitator of the process within the DOT organization.

TABLE OF CONTENTS

GUIDELINES FOR ENSURING AND MAXIMIZING THE QUALITY, OBJECTIVITY, UTILITY AND INTEGRITY OF INFORMATION DISSEMINATED BY THE DEPARTMENT OF TRANSPORTATION (DOT).

I.   WHAT IS THE BACKGROUND AND PURPOSE OF THESE GUIDELINES?

The Department of Transportation (DOT) is issuing guidelines to implement Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (PL 106-554). The Office of Management and Budget (OMB) previously issued Government-wide guidelines under Section 515, which direct Federal agencies subject to the Paperwork Reduction Act (44 U.S.C. Chapter 35) to establish and implement written guidelines to ensure and maximize the quality, utility, objectivity and integrity of the information that they disseminate. DOT's guidelines apply to a wide variety of substantive information dissemination activities in order to meet basic information quality standards set forth by Section 515. Under Section 515, the Department is responsible for carrying out the OMB information quality guidelines as well as for implementing its own guidelines that are set forth in this document. Consequently, when this document refers to "the guidelines," it should be taken to refer to the OMB guidelines, as applied to DOT programs and activities, as well as the DOT guidelines themselves, unless the context suggests otherwise.

The purpose of these guidelines is to provide a framework under which DOT will provide affected persons an opportunity to seek and obtain correction of information maintained and disseminated by DOT that does not comply with these guidelines. DOT has designated the Departmental Chief Information Officer (CIO) as the senior official responsible for DOT compliance with these guidelines. Heads of Departmental Organization are responsible for ensuring proper implementation of these Departmental guidelines. These final guidelines are effective October 1, 2002.

In implementing these guidelines, the Department acknowledges that ensuring the quality of information is an important management objective that takes its place alongside other Departmental objectives, such as ensuring the success of agency missions, observing budget and resource priorities and restraints, and providing useful information to the public. The Department intends to implement these guidelines in a way that will achieve all these objectives in a harmonious way.

II.   WHO ARE THE DOT ORGANIZATIONS TO WHICH THESE GUIDELINES APPLY?

These guidelines apply to the following DOT organizations.

Office of the Secretary (OST)
Bureau of Transportation Statistics (BTS)
Office of the Inspector General (OIG)

Federal Aviation Administration (FAA)
Federal Highway Administration (FHWA)
Federal Motor Carrier Safety Administration (FMCSA)
Federal Railroad Administration (FRA)
Federal Transit Administration (FTA)
Maritime Administration (MARAD)
National Highway Traffic Safety Administration (NHTSA)
Research and Special Programs Administration (RSPA)
Saint Lawrence Seaway Development Corporation (SLSDC)
Transportation Administrative Service Center (TASC)
Transportation Security Administration (TSA)
United States Coast Guard (USCG)

DOT organizations may adopt further guidance, consistent with these guidelines, applying to the specifics of their programs and information products.

III.   WHAT ARE THE SCOPE, NATURE AND LEGAL EFFECT OF THESE GUIDELINES?

a.   The Scope: These guidelines apply to certain information (both statistical and non-statistical) disseminated by DOT on or after October 1, 2002, regardless of when the information was first disseminated. Maintenance of information on DOT web sites or paper files does not, in itself, subject information disseminated before this date to the guidelines. However, the Department's policy is to treat as subject to the guidelines information that we maintain in a way that is readily available to the public and that continues to play a significant, active role in Department programs or in private sector decisions. These guidelines apply to all media (printed, electronic, or in other form). As is the intent of OMB's guidelines, DOT's guidelines will focus primarily on the dissemination of substantive information (e.g., reports, analyses, studies, summaries) rather than information pertaining to basic agency operations.

The standards of these guidelines apply not only to information that DOT generates, but also to information that other parties provide to DOT, if the other parties seek to have the Department rely upon or disseminate this information or the Department decides to do so. For example, suppose that a trade association, in commenting on a proposed rule, supplies a scientific or technical study or an economic analysis in support of its position on what the final rule should say. In order for DOT to rely on this information in a subsequent DOT dissemination of information (e.g., as part of the basis cited for decisions in the final rule), the quality of the trade association's information would have to be consistent with these guidelines. Likewise, if the Department disseminates information originally created by, for example, a contractor or consultant, these guidelines would apply.

12

The types of Departmental information not subject to these guidelines are outlined in Section IV.

b. <u>Nature and Legal Effect</u>: These guidelines are policy views of DOT. They are not intended to be, and should not be construed as, legally binding regulations or mandates. These guidelines are intended only to improve the internal management of DOT and do not create any right or benefit, substantive or procedural, enforceable at law or equity, by any party against the United States, its agencies (including the Department of Transportation or any DOT organization), officers, or employees, or any person.

IV. WHAT TYPES OF INFORMATION ARE NOT SUBJECT TO THESE GUIDELINES?

The following information is not subject to these guidelines:

a. Distribution of information that is intended to be limited to government employees, agency contractors or grantees;

b. Distribution of information that is intended to be limited to intra- or inter-agency use or sharing of government information; responses to requests under FOIA, Privacy Act, the Federal Advisory Committee Act or other similar laws;

c. Predisclosure Notification to Submitters of Confidential Commercial Information (FHWA Notice N1320.6);

d. Distribution limited to correspondence with individuals or persons (regardless of medium, e.g., electronic mail). The possibility of further distribution by the recipient does not cause such correspondence to be subject to these guidelines. However, information sent by letter to a wide variety of individuals (e.g., a "Dear Colleague" letter sent to heads of all recipients of financial assistance from a DOT operating administration) would be subject to coverage under these guidelines;

e. Archival records disseminated by Federal agency libraries, websites, or similar Federal data repositories; (e.g., inactive or historical materials in DOT libraries and other data collections - including bibliographies or responses to reference requests pertaining to such materials);

f. Public filings, such as material filed by DOT or non-DOT parties in DOT Dockets or by DOT in other agencies' dockets. For example, a study filed in the DOT docket by a commenter on a proposed rule does not become subject to these guidelines simply because it has been filed there. However, if the Department chooses to rely on the study in the rulemaking or another information product, the Department's study would become subject to these guidelines because of the Department's use of it;

g. Information intended to be limited to subpoenas and adjudicatory processes. For the purpose of these guidelines, these processes would include:

1. Court or administrative litigation (e.g., briefs and attachments, or other information that the Department submits to the court or other decision maker);

2. Administrative enforcement proceedings conducted by the Department;
3. Civil rights and personnel complaints and reviews conducted by the Department (e.g., under Titles VI and VII of the Civil Rights Act; the Americans with Disabilities Act; Sections 501, 504 and 508 of the Rehabilitation Act of 1973; Title IX of the Education Amendments of 1972, and Disadvantaged Business Enterprise matters);
4. Debarment and suspension matters, 49 CFR Part 29 (Federal-aid contracts) and 48 CFR Part 9 (direct contracts);
5. Merit System Protection Board matters (Sections 7511, 7543, and 70701 of Title 5, United States Code);
6. Matters before the Board of Correction of U.S.C.G. Military Records (Section 1552 of Title 10, United States Code and Part 52 of Title 33, Code of Federal Regulations (1996));
7. USCG Commandant decisions on Appeal and Review of Suspension and Revocation Proceedings, Sections 7702 and 7704, Title 46, United States Code)
8. Locomotive engineer certification matters, 49 CFR Part 240, Subpart E-Dispute Resolution procedures;

h. Hyperlinks to information that others disseminate (as well as paper-based information from other sources referenced but not adopted or endorsed by DOT);
i. Views or opinions, where the presenter makes it clear that what is being offered is someone's opinion rather than fact or the Department's views;
j. Information presented to Congress as part of the legislative or oversight processes (e.g., testimony of DOT officials, information or drafting assistance provided to Congress in connection with pending or proposed legislation or oversight) that has previously been subject to the guidelines, is primarily a statement of the views of the Department on an issue, or is provided to a member of Congress who then disseminates it publicly. However, the Department would treat studies or other factual information products that are presented to Congress, and that have not previously been subject to these guidelines, as being covered.
k. Press releases and other information of an ephemeral nature, advising the public of an event or activity of a finite duration - regardless of medium. Information products referenced in such releases may be subject to the guidelines, however (e.g., a study referred to in a press release); and
l. Procedural, operational, policy, and internal manuals prepared for the management and operations of DOT that are not primarily intended for public dissemination. This includes personnel notices such as vacancy announcements.

V. WHAT GENERAL STANDARDS OF QUALITY ARE DOT ORGANIZATIONS
IMPLEMENTING?

DOT has traditionally utilized standards, policies, and other operational guidelines
to ensure the quality of all its disseminated information. Incorporating these
guidelines further reinforces DOT's commitment of meeting higher standards of
quality prior to disseminating information to the public. The Department has made
implementation of these guidelines a part of its performance plan, including
performance goals and standards.

To ensure compliance with these guidelines, each DOT organization has appointed
a data quality official who will serve as the liaison for implementing these
guidelines within its organization. Each DOT organization is responsible for
modifying appropriate information technology and administrative policies to
comply with these guidelines.

OMB's guidelines define "quality" as an encompassing term comprising utility,
objectivity, and integrity. Therefore, the guidelines sometimes refer to these
statutory terms, collectively, as "quality." At a minimum, a basic standard of quality
will be ensured and established for all DOT information prior to its dissemination.
In addition, on-going disseminated information will be reviewed on a regular basis
to ensure all information is current and complies with these guidelines. Specifically,
DOT will set the following standards at levels appropriate to the nature and
timeliness of substantive information to be disseminated.

a. Utility: DOT organizations will assess the usefulness of the information to be
disseminated to the public. The originating office will continuously monitor the
information needs and develop new sources or revise existing methods, models, and
information products where appropriate;

b. Objectivity: DOT organizations will ensure disseminated information is
accurate, clear, complete, and unbiased, both as to substance and presentation, and
in a proper context. The originating office will use reliable data sources and sound
analytical techniques. To the extent possible and consistent with confidentiality
protections, the originating office will identify the source of disseminated
information so that the public can assess whether the information is objective.

The Department intends to follow a policy of determining, in consultation as
appropriate with relevant scientific and technical communities, when it is useful and
practicable to apply reproducibility standards to original and supporting data. In
making such determinations, the Department will be guided by commonly accepted
scientific, financial or statistical standards, as applicable. With respect to analytic
results, the Department's policies favor sufficient transparency about methods to
allow independent reanalysis by qualified members of the public. In situations
where public access will not occur (e.g., because of confidentiality requirements or
the use of proprietary models), the Department's policy is to apply and document

15

especially rigorous robustness checks. In any case, the Department's policy is to provide the maximum feasible transparency with respect to specific data sources, quantitative methods, and assumptions used.

In OMB's guidelines, one of the aspects of ensuring objectivity deals with the use of peer review. For information products to which peer review is relevant, OMB's guidelines create a rebuttable presumption that the information meets the OMB guidelines' objectivity standards if the data and analytic results have been subject to formal, independent, external peer review. Anyone seeking to rebut this presumption (i.e., as part of the request for correction process) would have the obligation of demonstrating that the information was not substantively accurate, clear, complete, or unbiased, both as to substance and presentation, and in a proper context.

With respect to *influential* scientific information disseminated by the DOT organizations regarding analysis of risks to human health, safety, and the environment, DOT organizations will adopt, with respect to the analysis in question, quality principles of the Safe Drinking Water Act of 1996 (42 U.S.C. 300g-1(b)(3)(A) & (B), except where the agency adapts these principles to fit the needs and character of the analysis.   These principles are as follows:

- Use the best available science and supporting studies conducted in accordance with sound and objective scientific practices, including peer-reviewed studies where available.

- Use data collected by accepted methods or best available methods (if the reliability of the method and the nature of the decision justifies the use of the data).

- In the dissemination of *influential* scientific information about risks, ensure that the presentation of information is comprehensive, informative, and understandable. In a document made available to the public, specify, to the extent practicable:
    - Each population addressed by any estimate of applicable effects.
    - The expected risk or central estimate of risk for the specific populations affected.
    - Each appropriate upper bound or lower-bound estimate of risk.
    - Each significant uncertainty identified in the process of the risk assessment and studies that would assist in reducing the uncertainty.
    - Any additional studies, including peer-reviewed studies, known to the agency that support, are directly relevant to, or fail to support the findings of the assessment and the methodology used to reconcile inconsistencies in the scientific data.

c. Integrity: DOT's policy is to ensure that information is protected from unauthorized access, corruption or revision (i.e., make certain disseminated information is not compromised through corruption or falsification). The Department is highly protective of information collected under pledges of confidentiality. To ensure integrity of information disseminated electronically, the departmental CIO has implemented an aggressive Information Technology Security Program (ITSP). This Program, which complies with the computer security provisions of the Paperwork Reduction Act covers all DOT information, data and resources collected, stored, processed, disseminated, or transmitted using DOT information systems, to include the physical facilities in which the information, data and resources are housed. The ITSP applies to all DOT employees, specifically those in positions of public trust, contractors, subcontractors and other users of DOT information technology and related resources. For example, Departmental policy, as stated in the Department's Information Resources Management Manual (DIRMM), states that all DOT web managers are responsible for establishing appropriate security safeguards for ensuring the "integrity" of the information disseminated on the web sites and complying with The Privacy Act of 1974, as amended, to ensure appropriate disclosure of information.

DOT is also subject to several other statutory requirements to protect the information it gathers and disseminates. These include, but are not limited to: The Paperwork Reduction Act of 1995, the Government Information Security Reform Act; OMB Circular A-130 (Management of Federal Information Resources, dated 12/12/85; revised 11/28/00), 49 U.S.C. 111(i) (BTS establishment) and the Trade Secret Act (18 U.S.C. 1905)

d. Accessibility: In 2001, the Departmental CIO issued a policy to ensure accessibility to all persons. This policy applies to all Departmental electronic and information technology developed, procured, maintained, or used by DOT organizations on or after June 21, 2001 unless covered by one of the following conditions: 1) micro-purchases made prior to January 1, 2003 (FAR 39.204(a); or 2) conformity would impose an undue burden on the agency (FAR 39.204(e) and 36 CFR 1194.2).

DOT's policy is to ensure that all disseminated information (including electronic and information technology media) is accessible to all persons (see DOT Section 508 Policy Statement).

VI. WHAT ADDITIONAL STANDARDS OF QUALITY ARE DOT ORGANIZATIONS IMPLEMENTING FOR STATISTICAL INFORMATION?

The 1991 Intermodal Surface Transportation Efficiency Act (ISTEA) created the Bureau of Transportation Statistics (BTS) within the Department of Transportation (DOT). Among other things, it made BTS responsible for: "issuing guidelines for the collection of information by the Department of Transportation required for

statistics … in order to ensure that such information is accurate, reliable, relevant, and in a form that permits systematic analysis." (49 U.S.C. 111 (c)(3))

A parallel requirement for developing guidelines emerged in the Paperwork Reduction Act of 1995. It tasked the Office of Management and Budget (OMB) to "develop and oversee the implementation of Government wide policy, principles, and guidelines concerning statistical collection procedures and methods; statistical data classification; statistical information presentation and dissemination; timely release of statistical data; and such statistical data sources as may be required for the administration of Federal programs." (44 U.S.C. 3504 (e)(3))

The Department's Bureau of Transportation Statistics (BTS) has established additional guidelines covering the Department's statistical programs. These guidelines (for statistical information) are based on structured planning, sound statistical methods and the principle of openness. Structured planning maintains the link between user needs and data system design. Sound statistical methods produce information (data and analysis results) that conforms to that design. Openness ensures that users of statistical information can easily access and interpret the information.

Each section of the statistical guidelines (Section VI a-e below) begins with a statement of principles, which contain definitions, assumptions, and rules or concepts governing action. The *principles* are followed by *guidelines,* which are specific recommended actions. Finally, each section concludes with *references and examples*.

To access DOT's Statistical Data Quality Guidelines click on section links below or refer to the Appendix of this report for a complete version of these guidelines.

a. Planning Data Systems
   - Data System Objectives
   - Data Requirements
   - Methods to Acquire Data
   - Sources of Data
   - Data Collection Design

b. Collection of Data
   - Data Collection Operations
   - Missing Data Avoidance

c. Processing Data
   - Data Editing and Coding
   - Handling Missing Data
   - Production of Estimates and Projections
   - Data Analysis and Interpretation

d. Dissemination of Information
- Publications and Disseminated Summaries of Data
- Micro data Releases
- Source and Accuracy Statements
- Pre-Dissemination Reviews

e. Evaluating Information Quality
- Data Quality Assessments
- Evaluation Studies
- Quality Control Systems
- Data Error Correction

VII. WHAT PROCESSES DOES DOT UTILIZE TO ENSURE INFORMATION QUALITY BEFORE IT IS DISSEMINATED?

DOT's policy is to conduct a pre-dissemination review on all information it disseminates on or after October 1, 2002. During this review, each DOT organization may utilize internal peer reviews and other review mechanisms to ensure the quality of all disseminated information. The costs and benefits of using a higher quality standard or a more extensive review process will be considered in deciding the appropriate level of review and documentation. With respect to information collection requirements covered by the PRA, the Department will ensure that these requirements are consistent with the guidelines and will so state in the PRA submission to OMB. The main components of DOT's pre-dissemination review policy are the following:

1. Allow adequate time for reviews, consistent with the level of standards required for the type of information to be disseminated. Consult with others (e.g., other DOT organizations, the public, State governments, etc…) that have a substantial interest in the proposed dissemination of the information.

2. Verify compliance with these guidelines (i.e., utility, objectivity, integrity and accessibility requirements) as well as other DOT organization specific guidance/procedures;

3. With respect to information a DOT organization believes to be "influential," maintain internal records of what additional standards will be applied to ensure its quality. Refer to Section XI of these guidelines for a discussion concerning how to think about the application of the term "influential" to DOT's information;

4. Ensure that the entire information product fulfills the intentions stated and that the conclusions are consistent with the evidence;

5. Indicate origin of data (when including data from an external source); and

6. Ensure that each program office can provide additional data on the subject matter of any covered information it disseminates.

VIII. WHAT ARE DOT'S PROCEDURES CONCERNING REQUESTS FOR CORRECTION OF INFORMATION?

a. May I request a correction of information from the Department?

If you are affected by information that the Department has disseminated on or after October 1, 2002 (i.e., if you are harmed because the information does not meet the standards of the guidelines or a correction of the information would benefit you), you may request that the Department correct that information. We regard information originally disseminated before October 1, 2002, as being subject to this correction process if it remains publicly available (e.g., it is posted on a DOT website or the Department makes it available on a generally distributed information source) and it continues to play a significant, active role in Department programs or in private sector decisions.

b. Where do I submit a request for correction of information?

You may make a request for correction of information or request for reconsideration via the Department's on-line correction request form (which can be accessed from the Department's Docket Management System (DMS) or the Department's customer support web site. (Access DOT's online request form here.) Although we prefer a completed on-line form, we will also respond to request in written form, by letter or fax, to the following address:

> U. S. Department of Transportation (DOT)
> Office of Dockets and Media Management
> SUBJECT: Request for Correction of Information
> Room PL-401
> 400 7th Street, S.W.,
> Washington, DC 20590
> Fax number: 202/366-7202

Note: If you are making a request for reconsideration, you should include a reference to the initially assigned DOT docket number. This will enable DMS personnel to correctly place your reconsideration request in the same docket as your initial request.

c. How does the Department process incoming requests for correction?

   We will post Incoming requests for correction, requests for reconsideration and DOT organizational responses on the DMS website. DMS will electronically notify designated data quality officials in DOT organizations that a request for correction/reconsideration is pending. In the event that DOT staff receive requests for correction/reconsideration by another means (e.g., mail to a DOT office), the staff will refer the request to the Office of Dockets and Media Management for inclusion in the DMS.

   You should be aware that the Department is not required to change, or in any way alter, the content or status of information simply based on the receipt of a request for correction. For example, DOT need not withdraw an information product from a website just because a request for correction has been received with respect to it. Nor does the receipt of a request, or its consideration of by the Department, result in staying or changing any action of the Department. The receipt of a request for correction likewise does not affect the finality of any decision of a DOT organization.

d. What Should You Include in a Request for Correction of Information?

   In keeping with the non-regulatory nature of these guidelines, this guidance-for the content of requests for correction of information - is not intended to constitute a set of legally binding requirements. However, DOT may be unable to process, in a timely fashion or at all, requests that omit one or more of the requested elements. DOT will attempt to contact and work with requesters to obtain additional information when warranted.

The following information is also provided on the electronic form mentioned in Section VIII (b) above.

1. You should include a statement that a request for correction of information is submitted under DOT's Information Dissemination Quality Guidelines;

2. You should include your name, mailing address, fax number, or e-mail address, telephone number and organizational affiliation, if any;

    Privacy Act Statement: DOT is authorized to obtain certain information under Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (Public Law No. 106-554, codified at 44 U.S.C. § 3516, note). Information (as identified in Section VI) will be needed to process requests and allow DOT to reply accordingly. This information is needed to respond to your request and initiate follow-up contact with you if required. Please do not send us your Social Security Number. You are advised that you do not have to furnish

the information but failure to do so may prevent your request from being processed. The information you furnish is almost never used for any purpose other than to process and respond to your request. However, we may disclose information to a congressional office in response to an inquiry made on your behalf, to the Department of Justice, a court, other tribunal when the information is relevant and necessary to litigation, or to a contractor or another Federal agency to help accomplish a function related to this process;

3. You should describe how the information in question affects you (e.g., how an alleged error harms you, and/or how the correction will benefit you);

4. You should clearly identify the report, data set, or other document that contains the information you want the Department to correct. Please be as specific as possible, include such identifying characteristics as title, date, how information was received (e.g., web-accessed);

5. You should clearly identify the specific information that you want the Department to correct. Please be as specific as possible, include such identifying characteristics as the name of DOT agency that originated the data, title, date, etc. For example, you should not rely solely on general statements that allege some type of error. Requests for information that are specific and provide evidence to support the need for correction will likely be more persuasive than requests that are general, unfocused, or simply indicate disagreement with the information in question;

6. You should specify, in detail, why you believe the information in question is inconsistent with the Department's and/or OMB's information quality guidelines (i.e., how the information fails to meet standards of integrity, utility, and/or objectivity);

7. You should specify your recommendations for what corrections DOT should make to the information in question and reasons for believing that these recommended corrections would make the information consistent with the DOT's and/or OMB's information quality guidelines;

8. You should include any documentary evidence you believe is relevant to your request (e.g., comparable data or research results on the same topic).

e. <u>May the Department reject a request for correction of information?</u>

Once the appropriate data quality official has received your request for correction of information from DMS, he/she will review your request and answer the following questions to determine if your request for correction is a

valid request:

1. Did DOT (as opposed to some other person or organization) actually disseminate the information you are requesting to be corrected?

2. Are you a person affected by the information in question?

3. Is the information about which you are requesting a correction from DOT covered by these Guidelines (see Section IV)?

4. Is your request frivolous or not germane to the substance of the information in question?

5. Has DOT responded previously to a request that is the same or substantively very similar? (Note: This does not mean that the Department would automatically reject a second or subsequent information correction request concerning the same information product. If one party made a request concerning one aspect of the information product, and a second party made a request concerning a different aspect of the same product, for example, our two requesters sought correction on substantively divergent grounds, it could be appropriate for the Department to consider both.).

6. With respect to information in a final rule, final environmental impact statement, or other final document on which there was an opportunity for public comment or participation with respect to the compliance of the information with these guidelines, could interested persons have requested the correction of the information at the proposed stage?

If the DOT organization determines that the answer to 1,2, or 3 is "no" or that the answer to Question 4, 5, or 6 is "yes," DOT has the discretion to reject your request without responding to it on its merits.

If DOT rejects your request on these grounds, the DOT organization will send a written response explaining why. Normally, the DOT organization will send this response within 60 calendar days of receiving your request. The DOT organization will file this response in the DMS.

If the DOT organization does not reject your request on these grounds, it will consider the request on its merits.

f. Who has the burden of proof with respect to corrections of information?

As the requester, you bear the burden of proof with respect to the necessity for correction as well as with respect to the type of correction you seek.

g. What determinations does the Department make concerning a request for correction of information?

If the DOT organization considers your request on its merits (that is, does not reject it for one of the reasons in paragraph (e) above), DOT will make the determination whether information subject to the DOT information quality guidelines complies with the guidelines. In doing so, the Department will consider whether the information or the request for correction is stale. If DOT did not disseminate this information recently (i.e., within one year of your request), or it does not have a continuing significant impact on DOT projects or policy decisions or on important private sector decisions, we may regard the information as stale for purposes of responding to a correction request, unless the complainant can show that he or she is affected by its dissemination. If we determine that information subject to the DOT information quality guidelines does not comply with the guidelines, the Department will decide what correction is appropriate to make in order to ensure compliance**. It should be noted that while the Department's policy is to correct existing information when necessary, the Department is not obligated to generate new or additional information to respond to requests for correction.**

Except with respect to information covered in VIII (e)(6) of these guidelines, the DOT organization provides a written response directly to the requester

The DOT organization will normally issue this response within-60 calendar days of receiving the request. If the DOT organization's response will take significantly longer than this period, the DOT organization will inform the requester that more time is required and indicate the reason why and an estimated decision date. This written explanation to the requester will also be filed in DMS.

h. How does the Department process requests for correction concerning information on which the Department has sought public comment?

Information in rulemakings and other documents concerning which public participation and comment are sought are subject to these guidelines. However, the Department may respond to requests for correction concerning such information through a different process than we use for other types of information. When the Department seeks public comment on a document and the information in it (e.g., a notice of proposed rulemaking (NPRM), studies cited in an NPRM, a regulatory evaluation or cost-benefit analysis pertaining to the NPRM; a draft environmental impact statement; a proposed policy notice or aviation order on which comment has been sought; a request for comments on an information collection subject to the Paperwork Reduction Act), there is an existing mechanism for responding to a request for correction. This mechanism is a final document that responds to public

comments (e.g., the preamble to a final rule).

Consequently, our response to a request for correction of such information will normally be incorporated in the next document we issue in the matter.

The Department would consider making an earlier response, if doing so (1) would not delay the issuance of the final action in the matter; and (2) the Department determined that there would be an unusually lengthy delay before the final document would be issued or the requester had persuaded the Department that there was a reasonable likelihood that the requester would suffer actual harm if the correction were not made before the final action was issued.

Once again, the DOT organization will place its response in the DMS. As noted above in VIII (d), a DOT organization may reject a request for correction with respect to information in a final document if there was an opportunity for public comment or participation with respect to the compliance of information to these guidelines and interested persons could have requested the correction of the information at the proposed stage.

IX.   HOW DO YOU SEEK RECONSIDERATION OF THE DEPARTMENT'S DECISION ON A REQUEST FOR CORRECTION?

a.   You may request reconsideration under this section if you have requested a correction of information under these guidelines, and you are not satisfied with the DOT organization's response. You should request reconsideration within 30 days of the date you received the DOT organization's decision on your original request for correction.

b.   You should send your request in the same manner, and to the same address, as provided in Section VIII of these guidelines. When completing the electronic DMS form, you need only complete Section II- "Request for Reconsideration".

c.   If there is an existing process for reconsidering a particular sort of information disseminated by DOT, the DOT organization will make use of that process. For example, if the information relates to a final rule a DOT organization has issued, and the DOT organization has an existing process for handling requests for the reconsideration of a final rule, the DOT organization would use that process. If the information relates to a final Environmental Impact Statement (EIS), the DOT organization may handle the request as though it were a request for a Supplemental EIS. If you state that you do not seek any change in the ultimate outcome of the process – such as a change in the content of the rule or the EIS – and the Departmental organization agrees that a correction is appropriate and can be made without adversely affecting that outcome, the Department will

respond to you as provided in paragraphs (d) - (f) of this section.

d.   In the absence of an existing applicable reconsideration process, the DOT organization will designate a reconsideration official. This official should be someone who can offer objectivity (i.e., was not involved in making the decision on the original request for correction or in producing the underlying information) and who has a reasonable knowledge of the subject matter. The official can either be within the DOT organization to which the request for reconsideration pertains or in another DOT organization. In the case of a request concerning influential information, the Department will designate a panel of officials to perform this function, and we may do so in other appropriate cases. Typically, such a panel would include one person from the DOT organization that made the initial determination and two from other DOT organizations.

e.   The reconsideration official will determine if additional corrective action is needed. This determination may pertain to the specific correction that is appropriate in a given case as well as to the issue of whether correction is merited at all. The reconsideration official will issue a written response to the requester stating the reasons for the decision.

f.   The DOT organization will normally issue this response within 60 calendar days of receiving the request for reconsideration. If the DOT organization's response will take significantly longer than this period, the DOT organization will inform the requester that more time is required and indicate the reason why and an estimated decision date. This response will be filed in the DMS. The reconsideration official's determination will also be filed in DMS.

X.   WHAT ARE THE DEPARTMENT'S REPORTING REQUIREMENTS

The Departmental Office of the Chief Information Officer will provide annual reports to OMB (which will include the number and nature of complaints received concerning agency compliance as well as how complaints were resolved) beginning January 1, 2004.

XI.   WHAT ARE THE DEFINITIONS ASSOCIATED WITH THESE GUIDELINES?

DOT has adopted the definitions of terms set forth in The Office of Management and Budget's Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies. The following information explains further the way that DOT uses some of these terms.

a. **Influential:** The following discussion is intended as guidance for DOT officials and other interested persons in determining whether scientific, financial, or statistical information is influential within the meaning of OMB's guidelines. This definition is important because it determines the level of scrutiny afforded to information.

The OMB guidelines define "influential" information as information that the agency reasonably can determine "will have or does have a clear and substantial impact on important public policies or important private sector decisions." The guidelines assign to DOT the task of defining this term in ways appropriate to the agency and its various programs.

The Department emphasizes that to be influential; information must have a clear and substantial impact. A clear and substantial impact, first of all, is one determined to have a high probability of occurring. If it is merely arguable that an impact will occur, or if it is a close judgment call, then the impact is probably not clear and substantial. It is necessary to be very sure before designating information as influential. The impact must be on "important" public policy or private sector decisions. Even if information has a clear and substantial impact, it is not influential if the impact is not on a public or private decision that is important to policy, economic, or other decisions.

OMB's guidelines' definition of this term applies only to scientific, financial, or statistical information. The definition does not address other types of information, no matter how important the information may seem to be. It should also be noted that the definition applies to "information" itself, not to decisions that the information may support. Even if a decision or action by DOT is itself very important, a particular piece of information supporting it may or may not be "influential."

In rulemaking, influential information is scientific, financial, or statistical information that can reasonably be regarded as being one of the major factors in the resolution of one or more key issues in a significant rulemaking, as that term is defined in Executive Order 12866. This part of this standard reflects the "clear and substantial impact" language in the OMB guidelines language. The reference to key issues on significant rules reflects the "important" public policy language of the guidelines.

In non-rulemaking contexts, DOT will consider two factors – breadth and intensity – in determining whether information is influential.

Every decision DOT makes based on disseminated information is important to someone. That does not mean that disseminated information used for each decision is influential, as the term is used in the guidelines.

In determining whether information is influential, DOT organizations should consider whether the information affects a broad range of parties. Information that affects a broad, rather than a narrow, range of parties (e.g., an entire industry or a significant part of an industry, as opposed to a single company) is more likely to be influential.

Departmental organizations will also consider whether the information has an intense impact. Information that has a low cost or modest impact on affected parties is less likely to be influential than information that can have a very costly or crucial impact. In considering whether information has a high-intensity impact, DOT organizations would generally find that the same kinds of factors that cause a rule to be an economically significant rule, as defined in Executive Order 12866, would lead us to judge the impact of information in non-rulemaking context to be intense. Even in the absence of these factors, however, we could determine that information has an intense impact. This determination calls for the application of sound case-by-case judgment by officials knowledgeable of the issues and parties involved.

Information that has an intense impact on a broad range of parties would be regarded as influential. Information that affects a broad range of parties, with a low-intensity impact, or information that affects a narrow range of parties, with a high-intensity impact, may or may not be influential. In making this determination, the Department would also take into account the overall magnitude of the impact of the information, not only its impact on a per capita or per unit basis.

Departmental organizations may designate certain classes of information as "influential" or not in the context of their specific programs. Absent such designations, DOT organizations will determine whether information is influential on a case-by-case basis, using the principles articulated in these guidelines.

The "influential" designation is intended to be applied to information only when clearly appropriate. DOT organizations should not designate information products or types of information as influential on a regular or routine basis. Nor should DOT organizations actually place an "influential" label in the title page or text of an information product.

b. **Reproducibility.** Documented methods are capable of being used on the same data set to achieve a consistent result. For more information on this term, please refer to OMB's guidelines.

c. **Dissemination.** As provided in OMB's guidelines, these guidelines apply only to information disseminated on or after October 1, 2002. The fact that an information product that was disseminated by DOT before this date is still maintained by the Department (e.g., in DOT's files, in publications that DOT

continues to distribute on a website) does not make the information subject to these guidelines or to the request for correction process. As noted above, the Department's policy is to treat as subject to the guidelines information that we maintain in a way that is readily available to the public and that continues to play a significant, active role in Department programs or in private sector decisions.

For example, suppose that DOT first issued a study in 1999. The study is relied upon in a 2000 DOT organization publication, and the DOT organization makes the publication available on its website. This study is not subject to these guidelines or to the request for correction process just because it is "archived" in an available paper publication or website. However, if DOT issues a notice of proposed rulemaking in 2003 that relies on the same study, then it becomes subject to these guidelines - because it then has been disseminated (or, one might say "re-disseminated") after October 1, 2002.

d. **Departmental organizations.** Offices within the Office of the Secretary, Operating Administrations (OA), offices, divisions, and comparable elements of the DOT.

e. **Departmental Chief Information Officer (CIO).** The Departmental CIO is the senior management official responsible for the DOT Information Dissemination Quality Program.

f. **Data Quality Administrator (DQA)**. Designated representative in the Office of the CIO (S-80) responsible for compiling agency reports and serving as agency liaison to OMB.

g. **Data Quality Official (DQO)**. The DQO serve as the point of contact for the Departmental CIO/Data Quality Administrator and will be responsible for implementing these guidelines within their organization.

h. **Docket Management System (DMS).** DMS is an electronic, image-based database in which all DOT docketed information is stored for easy research, and retrieval.

i. **Docket.** A docket is an official public record. DOT publishes and stores on-line information about proposed and final regulations, copies of public comments on proposed rules, and related information in the DMS. DOT uses this docketed material when making regulatory and adjudicatory decisions, and makes docketed material available for review by interested parties. Specific documents covering the same issues are stored together in a docket.

j. **Transparency:** Includes both presentation and the reporting of information sources and limitations.

1. Introduction

Quality of data has many faces. Primarily, it has to be *relevant* (i.e., useful) to its users. Relevance is achieved through a series of steps starting with a planning process that links user needs to data requirements. It continues through acquisition of data that is *accurate* in measuring what it was designed to measure and produced in a *timely* manner. Finally, the data must be made *accessible* and *easy to interpret* for the users. In a more global sense, data systems also need to be *complete* and *comparable* (to both other data systems and to earlier versions). The creation of data that address all of the facets of quality is a unified effort of all of the development phases from the initial data system objectives, through system design, collection, processing, and dissemination to the users. These sequential phases are like links in a chain. The sufficiency of each phase must be maintained to achieve relevance. This document is intended to help management and data system "owners" achieve relevance through that sequential process.

1.1   Legislative Background

The 1991 Intermodal Surface Transportation Efficiency Act (ISTEA) created the Bureau of Transportation Statistics (BTS) within the Department of Transportation (DOT). Among other things, it made BTS responsible for: "issuing guidelines for the collection of information by the Department of Transportation required for statistics … in order to ensure that such information is accurate, reliable, relevant, and in a form that permits systematic analysis." (49 U.S.C. 111 (c)(3))

A parallel requirement for developing guidelines emerged in the Paperwork Reduction Act of 1995. It tasked the Office of Management and Budget (OMB) to "develop and oversee the implementation of Government wide policy, principles, and guidelines concerning statistical collection procedures and methods; statistical data classification; statistical information presentation and dissemination; timely release of statistical data; and such statistical data sources as may be required for the administration of federal programs." (44 U.S.C. 3504 (e)(3))

Lastly, the Consolidated Appropriations Act of 2001, section 515, elaborated on the Paperwork Reduction Act, requiring OMB to issue guidelines ensuring the quality of disseminated information by 9/30/2001 and each federal agency to issue guidelines by 9/30/2002.

1.2   OMB Guidelines for Ensuring Information Quality

On 28 September 2001, OMB published a notice in the Federal Register (finalized as 67 FR 8452, February 22, 2002) that required agencies to issue guidelines for

ensuring and maximizing the quality of information disseminated by federal agencies.

As defined in the OMB guidance, quality consists of:
- Utility, i.e., the usefulness of information to intended users,
- Objectivity in presentation and in substance, and
- Integrity, i.e., the protection of information from unauthorized access or revision.

Agencies were required to develop guidelines, covering all information disseminated on or after October 1, 2002, regardless of format. Agencies were also required to develop a process for pre-dissemination review of information, an administrative mechanism allowing the public to request correction of information not complying with the guidelines, and an annual report to OMB indicating how the public requests were handled by the mechanism.

These guidelines incorporate the statistical aspects of the OMB guidelines as a baseline and elaborate on its recommendations to produce statistical guidelines adapted for the Department of Transportation.

1.3 Applicability

These guidelines apply to all statistical information that is disseminated on or after 1 October 2002 by agencies of the Department of Transportation (DOT) to the public using the "dissemination" definition in the OMB guidelines. That definition exempts a number of classes of information from these guidelines. Major types of exempted information are listed below. A more detailed list is provided in section IV of the DOT Information Dissemination Quality Guidelines, of which this document is a subsection.
- Information disseminated to a limited group of people and not to the public in general.
- Archival records that are inherently not "active."
- Materials that are part of an ad judicatory process.
- Hyperlinked information.
- Opinion offered by DOT staff in professional journals.

DOT disseminated data contain a lot of information provided by "third party sources" like the states, industry organizations, and other federal agencies. These guidelines apply to that disseminated data unless exempted for other reasons discussed above. However, DOT guidelines indicating design, collection, and processing methods do not apply to data acquisition steps performed by non-federal sources. Steps performed by federal sources outside DOT before providing the data to DOT will be governed by the agency's own guidelines in accordance with this legislation. For data provided to DOT by third party sources, these guidelines primarily emphasize disseminating information about

data quality, the DOT processing methods, and analysis of the data provided to the users.

1.4 Types of DOT Statistical Data Collected

The recommendations within these guidelines apply to a wide range of data collection types. They include reporting collections, surveys, and special studies.

Reporting collections are set up to be automatic delivery of data into the data system. They collect incident information from government (federal, state, or local) and industry sources and periodic information on transportation flow and volume from government and industry. Incident data tend to cover all incidents (e.g., fatal accidents), though some data may be sampled due to its sheer volume (e.g., highway injuries). Flow and volume collections are a mixture of 100% collection and sampled data. Surveys and special studies are more of an outreach form of data collection. Surveys and studies are usually conducted using some form of sampling.

Samples taken for any data collection may be selections of people or organizations from lists, samples of geographic areas or sections of highway, or samples of time segments.

1.5 Overview of the Statistical Guidelines

The quality guidelines for statistical information are based on structured planning (section 2), sound statistical methods (sections 3 and 4) and the principle of openness (sections 5 and 6). Structured planning maintains the link between user needs and data system design. Sound statistical methods produce information (data and analysis results) that conforms to that design. Openness ensures that users of statistical information can easily access and interpret the information.

Each section begins with a statement of *principles*, which contain definitions, assumptions, and rules or concepts governing action. The *principles* are followed by *guidelines*, which are specific recommended actions with examples. Finally, each section concludes with *references*.

1.6 Statistical Guidelines Relationship to DOT's Information Dissemination Quality Guidelines

These statistical guidelines are a subset of the DOT Information Dissemination Quality Guidelines. Chapters 2 through 6 discussed above form section VI, paragraphs a – e in that document.

2. Planning Data Systems

Data systems produced within a DOT agency are created to fulfill user needs. Users can be those within DOT and outside. The data are compiled to satisfy an external user need, measure success toward a strategic goal (internal user), or used as a tool necessary to perform work toward a goal (internal user). Data system planning consists of four stages: collecting user needs, development of objectives for the system, translation of those objectives into data requirements, and planning of the top-level methods that will be used to acquire the data.
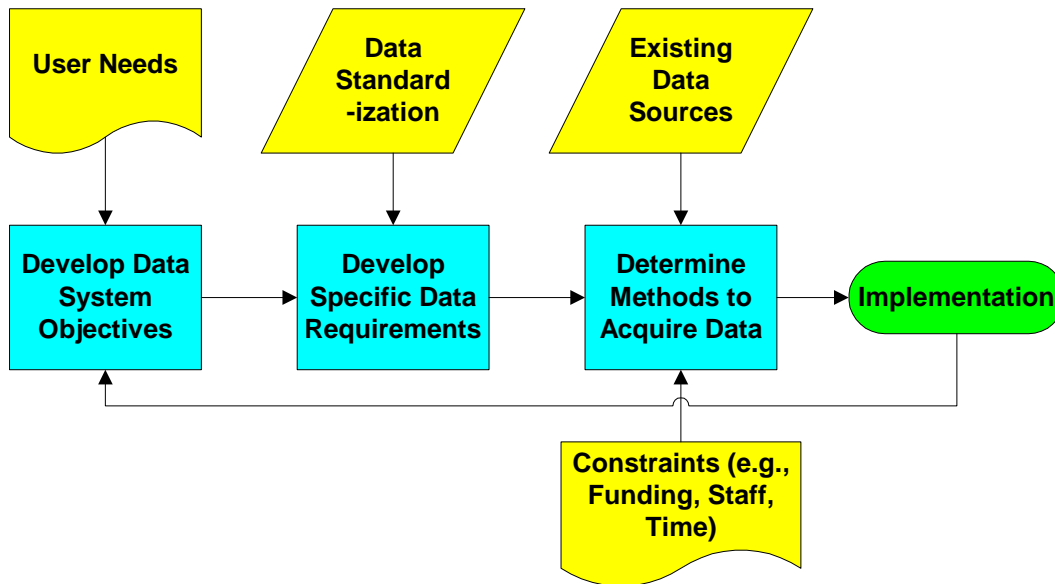
2.1 Data System Objectives

Principles

- A "data system" is any collection of information that is used as a source by any Government entity to disseminate information to the public, along with the planning, collection, processing, and evaluation. A data system can cover any combination of information treated as a single system for the sake of documentation and other guideline issues.

- The "system owner" as used in these guidelines is the organizational entity whose strategic plan and budget will guide the creation or continued maintenance of the data system.

- "Users" of a data system are people or organizations who use information products that incorporate data from the system, either in raw form or in statistics. "Major Users" of the data system are system users identified as such in strategic plans and legislation supporting the creation and maintenance of the data system. "User needs" should be in the form of questions that specific users want to be answered.

- "Objectives" of the data system describe what federal programs and external users will accomplish with the information.

- System objectives in clear, specific terms, identifying data users and key questions to be answered by the data system, will help guide the system development to produce the results required.

- Just as user needs change over time, the objectives of the data system will need to change over time to meet new requirements.

- Users will benefit from knowing the objectives that guided the system design.

Guidelines

- Data system objectives should be written in terms of the questions that need to be answered by the data; not in terms of the data itself.

- Every data system objective should be traceable to user needs.

> For example, NHTSA, as an internal user of the Fatality Analysis Reporting System (FARS) has a primary goal to improve traffic safety and a need for information related to that goal. So, one objective for the Fatality Analysis Reporting System (FARS) could be to provide an overall measure of highway safety to evaluate the effectiveness of highway safety improvement efforts.

- The system owner should develop and update the data system objectives in partnership with critical users and stakeholders. The owner should have a process to regularly update the system as user needs change.

> For example, for the Highway Performance Monitoring System (HPMS), one of the objectives may be: to provide state and national level measures of the overall condition of the nation's public roads for Congress, condition and performance information for the traveling public, and information necessary to make equitable apportionments of highway funds to the states. The specific needs of major users have to be monitored and continuously updated.
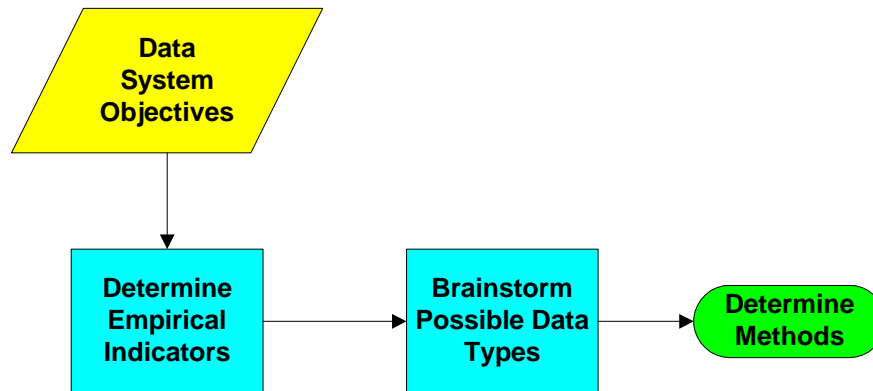
- Objectives should include timeliness of the data related to user needs.

- The current data system objectives should be documented and made available to the public, unless restricted.

- The updating process should be documented and include how user
  information is collected.

References

- Huang, K., Y.W. Lee, and R.Y. Wang. 1999. Quality Information and
  Knowledge. Saddle River, NJ: Prentice Hall.

2.2 Data Requirements

```
        ┌──────────────┐
       /  Data          /
      /   System       /
     /    Objectives  /
    └──────────────┘
            │
            ▼
  ┌──────────────┐    ┌──────────────┐    ┌──────────────┐
  │  Determine   │    │  Brainstorm  │    │  Determine   │
  │  Empirical   │──▶ │ Possible Data│──▶ │  Methods     │
  │  Indicators  │    │   Types      │    │              │
  └──────────────┘    └──────────────┘    └──────────────┘
```

Principles

- An "empirical indicator" is a characteristic of people, businesses, objects, or
  events (e.g., people or businesses in a city or state, cars or trains in the United
  States, actions at airports, incidents on highways).

    Examples: The level of success in stopping illicit drug smuggling
    into the U.S. over maritime routes. The level of use of public
    transit in a metropolitan area.

- Before deciding on what data should be in a data system or how to acquire
  them, the data system objectives need to be linked to more specific "empirical
  indicator," from which data requirements will be derived.

    > Example: For FARS, the objective "To provide an overall
    > measure of highway safety" leads to an empirical indicator of
    > "Injury or death of people on the highways of the U.S."

- Empirical indicators related to objectives can be outcomes that change as
  objectives are achieved, outputs from agency accomplishments related to an
  objective, efficiency concepts, inputs, and quality of work.

- From the empirical indicators, data requirements are created for possible measurement of each empirical indicator.

- Maintaining the link from data system objectives to empirical indicators to data requirements will help to ensure "relevance" of the data to users.

- In the data requirements, the use of standard names, variables, numerical units, codes, and definitions allow data comparisons across databases.

- Besides data that are directly related to strategic plans, additional data may be required for possible cause and effect analysis.

> For example, data collected for traffic crashes may include weather data for causal analysis.

Guidelines

- Each data system objective should have one or more "indicators" that need to be measured. Characteristics or attributes of the target group that are the focus of the objective should be covered by one or more empirical indicators.

  For HPMS, the objective "to provide a measure of highway road use" can lead to the empirical indicator of "the annual vehicle miles of travel on the interstate system & other principle arteries."

- The empirical indicators should be those characteristics which, when changing in a favorable way, indicate progress toward achievement of an objective.

> Note: Exceptions to this description are measures of magnitude, such as a total population or total vehicle miles traveled. These are "denominator measures" used to allow comparisons over time.

- Once the empirical indicators are chosen, develop data requirements needed to quantify them.

  Example: For HPMS, the empirical indicator, "the annual vehicle miles of travel on the interstate system & other principle arteries" can lead to a data requirement for state-level measures of annual vehicle-miles traveled accurate to within 10 percent at 80 percent confidence.

- There is usually more than one way to quantify an empirical indicator. All reasonable measures should be considered without regard to source or availability of data. The final data choices will be made in the "methods"

phase based on ease of acquisition, constraining factors (e.g., cost, time, legal factors), and accuracy of available data.

Example: A concept of commercial airline travel "delay" can be measured as a percent of flights on-time in accordance with schedule, or a measure of average time a passenger must be in the airport including check in, security, and flight delay (feasibility of measure is not considered at this stage).

- In the data requirements, each type of data should described in detail. Key variables should include requirements for accuracy, timeliness, and completeness. The accuracy should be based on how the measure will be used for decision-making.

  > Example: For FARS, the concept, "The safety of people and pedestrians on the highways of the U.S." can lead to data requirements for counts of fatalities, injuries, and motor vehicle crashes on U.S. highways and streets. The fatalities for a fiscal year should be as accurate as possible (100% data collection), available within three months after the end of the fiscal year, and as complete as possible. The injury counts in traffic crashes for the fiscal year totals should have a standard error of no more than 6 percent, be available within three months after the end of the fiscal year, and have an accident coverage rate of at least 90 percent.

- When selecting possible data, consider standardization with other databases. First, consider measures used for similar concepts in other DOT databases. Second, consider measures for similar concepts in databases outside DOT (e.g., The Census). Coding standards should be used where coding is used and made part of the data requirements. Such standardization leads to "coherence" across datasets.

  > Examples: the North American Industry Classification System (NAICS) codes, the Federal Information Processing Standards (FIPS) for geographic codes (country, state, county, etc.), the Standard Occupation Codes (SOC), International Organization for Standardization (ISO) codes (money, countries, containers)

- The current data system empirical indicators and data requirements should be documented and clearly posted with the data.

References

- The Federal Information Processing Standards (FIPS) home page, http://www.itl.nist.gov/fipspubs/.

- The North American Industry Classification System,
  http://www.census.gov/epcd/www/naics.html.
- OMB Primer on Performance Management dated 2/28/1995.
- American Association for Public Opinion Research. 1998. "Standard
  Definitions – Final Dispositions of Case Codes and Outcome Codes for RDD
  Telephone Surveys and In-Person Household Surveys,"
  http://www.aapor.org/ethics/stddef.html.

2.3 Methods to Acquire Data

Given data requirements for a wide range of possible measures, the next phase is to consider the realities associated with gathering the data to construct estimates and perform analysis. After looking at the ease of data acquisition, complexity of possible acquisition approaches, budget restrictions, and time considerations, the list of possible measures is likely to be reduced to a more reasonable level. First, consider possible sources of data and then the process of acquiring it.
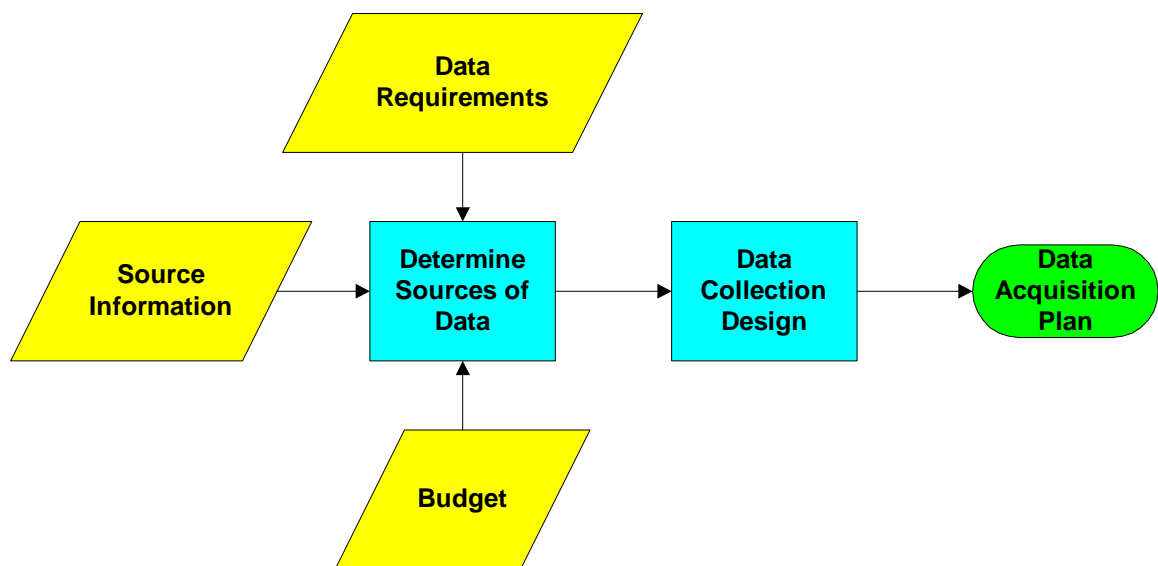
The more critical data needs invariably require greater accuracy. This in turn usually leads to a more complex data collection process. As the process gets more complex, there is no substitute for expertise. If the expertise for a complex design is not available in-house, consider acquiring the expertise by either contacting an agency specializing in statistical data collection like the Bureau of Transportation Statistics or by getting contractual support.

2.4 Sources of Data

Principles

- A common arrangement in transportation is a reporting collection in which the target group automatically sends data. Most of these are dictated by law or regulation. That limits the collection planning to working out the physical details.

  > For example: 46 USC Chapter 61 specifies a marine casualty reporting collection, while 46 CFR 4.05 specifies details.

- If existing data can be found that addresses data requirements, it is by far the most efficient (i.e., cheapest) approach to data acquisition. Sources of existing data can be current data systems or administrative records.

- "Administrative records" are data that are created by government agencies to perform facilitative functions, but do not directly document the performance of mission functions (National Archives definition). In addition to providing a source for the data itself, administrative records may also provide information helpful in the design of the data collection process (e.g., sampling lists, stratification information).

  > For example, state driver's license records, social security records, IRS records, boat registration records, mariner license records.

- Another method, less costly than developing a new data collection system, is to use existing data collections tailored to your needs. The owner of such a system may be willing to add additional data collection or otherwise alter the collection process to gather data that will meet data requirements.

  > For example, the Bureau of Transportation Omnibus survey is a monthly transportation survey that will add questions related to transportation for special collections of data from several thousand households. This method could be used if this process is accurate enough for the data system needs.

- The "target group" is the group of all people, businesses, objects, or events about which information is required.

  > For example, the following could be target groups: all active natural gas pipelines in the U.S. on a specific day, traffic crashes in FY2000 involving large trucks, empty seat-miles on the MARTA rail network in Atlanta on a given day, hazardous material

> incidents involving radioactive material in FY2001, mariners in distress on a given day, and U.S. automobile drivers.

- One possible approach is to go directly to the "target group," either all of them (100%) or a sample of them. This would work with people or businesses.

- Another method frequently necessary with transportation data is the use of third party sources. Third party sources are people, businesses, or even government entities that have knowledge about the target group or collect information for other purposes, such as investigators, observers, or service providers (e.g., doctors).

  > Examples: traffic observers, police observers, investigators, bus drivers counting passengers, state data collectors.

Guidelines

- Research whether government and private data collections already have data that meet the data requirements. Consider surveys, reporting collections, and administrative records.

- If existing data meet some but not all of the data requirements, determine whether the existing data systems can be altered to meet the data needs.

  > For example, another agency may be willing to add to or alter their process in exchange for financial support.

- A primary consideration in whether to gather data from the target group or an indirect source is access to the group; all of them. A 100% data gathering would obviously need access to the entire target group. A sample approach will not include the entire target group, but all members should have a non-zero (and known) probability of selection, or the sampling will not necessarily be representative of the target group.

- Consider getting information directly from the target group (if they are people or businesses), having the target group observed (events as they occur), or getting information about the target group from another source (third party source discussed above).

- In some situations, the information desired is not directly available. In this case, consider collecting related information that can be used to derive or estimate the information required.

> For example: Collecting the number of people on and off a bus at each stop combined with a separate estimate of trip length between stops to estimate passenger miles.

- When using third-party data for a data system, ensure that the data from the third party meets data requirements. If the third party source is mandated or a sole source for the data, gather information on each data requirement, as available.

- The choices made for sources and their connection to the data requirements should be documented and clearly posted with the data, or with disseminated output from the data.

References

- Electronic Records Work Group Report to the National Archives and Records Administration dated September 14, 1998.

2.5 Data Collection Design

Principles

- The design of data collection is one of the most critical phases in developing a data system. The accuracy of the data and of estimates derived from the data are heavily dependent upon the design of data collection.

> For example, the accuracy is dependent upon proper sample design, making use of sampling complexity to minimize variance. The data collection process itself will also determine the accuracy and completeness of the raw data.

- Data collection from 100% of the target group is usually the most accurate approach, but is not always feasible due to cost, time, and other resource restrictions. It also is often far more accurate than the data requirements demand and can be a waste of resources.

- A "probability sample" is an efficient way to automatically select a data source representative of the target group with the accuracy determined by the size of the sample.

- When sampling people, businesses, and/or things, sampling lists (also known as frames) of the target group are required to select the sample. Availability of such lists is often a restriction to the method used in data collection.

- For most statistical situations, it is usually important to be able to estimate the variance along with estimating the mean or total.

- Sample designs should be based on established sampling theory, making use of multi-staging, stratification, and clustering to enhance efficiency and accuracy.

- Sample sizes should be determined based on the data requirements for key data, taking into account the sample design and missing data.

Guidelines

- The data collection designer should use a probability sample, unless a 100% collection is required by law, necessitated by accuracy requirements, or turns out to be inexpensive (e.g., data readily available).

     For example, a system that collects data to estimate the total vehicle miles traveled (VMT) for a state of the U.S. cannot possibly collect 100 percent of all trips on every road, so a sampling approach is necessary. However, when it comes to collecting passenger miles for a large transit system, it may be possible with fare cards and computer networks to collect 100% of passenger miles.

- The sample design should give all members of the target group a non-zero (and known) probability of being represented in the sample.

     DANGER => Samples of convenience, such as collecting transportation counts at an opportune location, will produce data, but it will almost always be biased. Whereas, randomly selecting counting sites from all possible locations will be statistically sound (with allowances due to correlations between locations).

- The design of any samples should be based on established sampling theory. Determine sample size using appropriate formulas to ensure data requirements for accuracy are met with adjustments for sample design and missing data. Use an appropriate random method to select sample according to the design.

- If some form of sampling is used, design the data collection to collect sufficient information to estimate the variance of each estimate to be produced.

- The collection design and its connection to the data requirements should be documented and clearly posted with the data, or with disseminated output

from the data. The documentation should include references for the sampling theory used.

- If the data collection process performed by DOT uses sampling, a statistician or other sampling expert should develop or review the design.

- If the data system uses third party data collected using sampling, sample design information should be collected and provided with collection design documentation, when available

References

- Cochran, William G., *Sampling Techniques* (3rd Ed.), New York: Wiley, 1977.

3. Collection of Data

Given the collection design, the next phase in the data acquisition process is the collection process itself. This collection process can be a one-time execution of a survey, a monthly (or other periodic) data collection, a continuous reporting of incident data, or a compilation of data already collected by one or more third parties. The physical details of carrying out the collection are critical to making the collection design a reality.
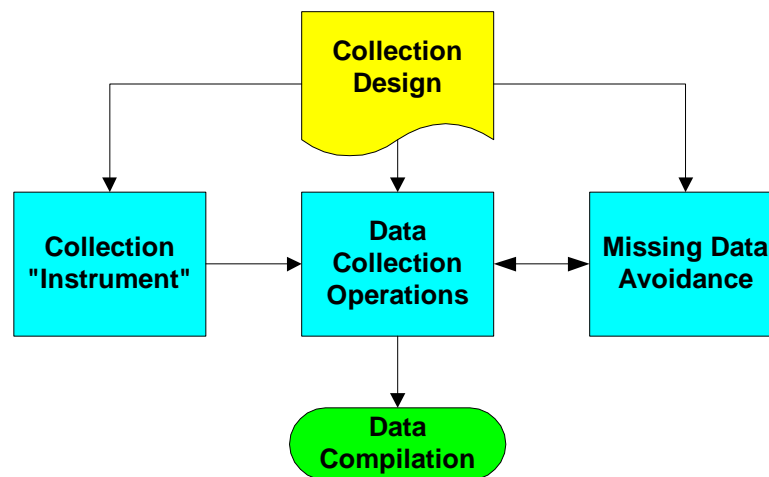
3.1 Data Collection Operations

Principles

- The collection "instruments" are forms, questionnaires, automated collection screens, and file layouts used to collect the data. They consist of sets of questions or annotated blanks on paper or computer that request information from data suppliers. They should be designed to maximize communication to the data supplier.

- Data collection includes all the processes involved in carrying out the data collection design to acquire data. Data collection operations can have a high impact on the ultimate data quality.

- The data collection method should be appropriate to the data complexity, collection size, data requirements, and amount of time available.

   Examples: A reporting collection will rely partially on the required reporting process, but will also follow-up for missing data. Similarly, a large survey requiring a high response rate will often start off with a mail out, followed by telephone contact, and finally by a personal visit.

- Specific data collection environmental choices can significantly affect error introduced at the collection stage.

  For example, if the data collector is collecting as a collateral duty or is working in a uncomfortable environment, it may adversely affect the quality of the data collected. Also, if the data are particularly difficult to collect, it will affect the data quality.

- Conversion of data on paper to electronic form (e.g., key entry, scanning) introduces a certain amount of error which must be controlled.

- Third party sources of data may introduce some degree of error in their collection processes.

Guidelines

- Collection instruments are clearly defined for data suppliers, with entries in a logical sequence, reasonable visual cues, and limited skip patterns. Instructions should help minimize missing data and response error.

- A status tracking procedure should be used to ensure that data are not lost in mailings, file transfers, or collection handling. A tracking system for incoming third-party data should ensure that all required data are received.

- Data entry of paper forms should have a verification process ensuring that data entry errors remain below set limits based on data accuracy requirements.

  For example, the verification samples of key entry forms can be based on an average outgoing quality limit for batches of forms. A somewhat more expensive approach would be 100 percent verification.

- Make the data collection as easy as possible for the collector.

- If interviewers or observers are used, a formal training process should be established to ensure proper procedures are followed.

- Data calculations and conversions at the collection level should be minimized.

  For example, if a bus driver is counting passengers, they should not be doing calculations such as summations. The driver should record the raw counts and calculations should be performed where they are less likely to result in mistakes.

- The collection operation procedures should be documented and clearly posted with the data, or with disseminated output from the data. If third party data collection is used, procedures used by the third party should be provided as well.

References

- Federal Committee on Statistical Methodology. 1983. *Approaches to Developing Questionnaires.* Washington, DC: U.S. Office of Management and Budget (Statistical Policy Working Paper 10).
- Groves, R. 1989. *Survey Errors and Survey Costs.* New York, NY: Wiley, Chs. 10 & 11.

3.2   Missing Data Avoidance

Principles

- Some missing data occur in almost any data collection effort. Unit-level missing data occur when a report that should have been received is completely missing or is received and cannot be used (e.g., garbled data, missing key variables). Item-level missing data occur when data are missing for one or more items in an otherwise complete report.

> For example, for an incident report for a hazardous material spill, unit-level missing data occur if the report was never sent in. It would also occur if it was sent in, but all entries were obliterated. Item-level missing data would occur if the report was complete, except it did not indicate the quantity spilled.

- The extent of unit-level missing data can sometimes be difficult to determine. If a report should be sent in whenever a certain kind of incident occurs, then non-reporters can only be identified if crosschecked with other data sources. On the other hand, if companies are required to send in periodic reports, the previous period may provide a list of the expected reporters for the current period.

> Both can also be true for item-level missing data. For example, in a travel survey asking for trips made, forgotten trips would not necessarily be known.

- Some form of missing data follow-up will dramatically reduce the incidents of both unit-level and item-level missing data.

> For example, a process to recontact the data source can be used, especially when critical data are left out. A series of recontacts may be used for unit nonresponse. Incident reporting collections can use some form of cross-check with other data sources to detect when incidents occur, but are not reported.

- When data are supplied by a third-party data collector, some initial data check and follow-up for missing data will dramatically reduce the incidents of missing data.

Guidelines

- All data collection programs should have some follow-up of missing reports and data items, even if the data are provided by third-party sources.

> For example, for surveys and periodic reports, it is easy to tell what is missing at any stage and institute some form of contact (e.g., mail out, telephone contact, or personal visit) to fill in the missing data. For incident reports, it is a little more difficult, as a missing report may not be obvious.

- For incident reporting collections where missing reports may not be easily tracked, some form of checking process should exist to reduce missing reports.

- For missing data items the data collection owner should distinguish between: critical items like items legally required or otherwise important items (e.g., items used to measure DOT or agency performance).

- The missing data avoidance procedures should be documented and clearly posted with the data, or with disseminated output from the data.

- Data collection program design documentation should address how the collection process was designed to produce high rates of response.

- If data is collected by a third party, the data collection program documentation should indicate how the third party deals with missing data, if that documentation is available.

References

- Groves, R.M. and M.P. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York, NY: Wiley.

4. Processing Data

Once the collected data is in electronic form, some "processing" is usually necessary to mitigate obvious errors, and some analysis is usually necessary to convert data into useful information for decision documents, publications, and postings for the Internet.
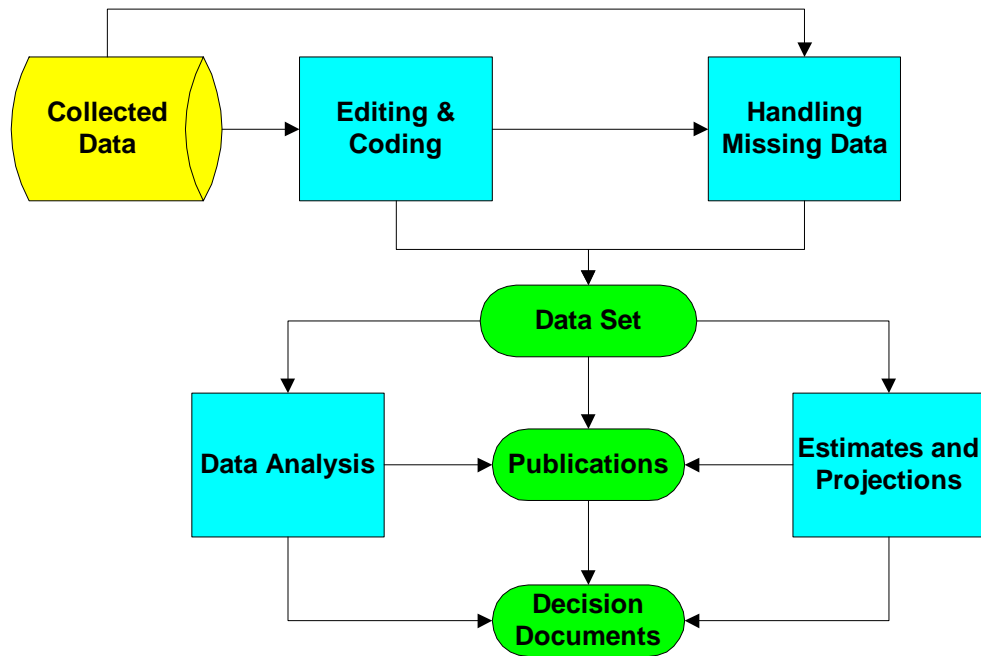
4.1 Data Editing and Coding

Principles

- Data editing is the application of checks that identify missing, invalid, duplicate, inconsistent entries, or otherwise point to data records that are potentially in error.

- Typical data editing includes range checks, validity checks, consistency checks (comparing answers to related questions), and checks for duplicate records.

- For numerical data, "outliers" are not necessarily bad data. They should be examined for possible correction, rather than systematically deleted.

    Note: By "examine" we mean you can check the original forms, compare data items with each other for consistency, and/or follow-up with the original source, all to see if the data are accurate or error has been introduced.

- Editing is a final inspection-correction method. It is almost always necessary, but data quality is better achieved much earlier in the process through clarity of definitions, forms design, data collection procedures, etc.

- Coding is the process of adding codes to the data set as additional information or converting existing information into a more useful form. Some codes indicate information about the collection. Other codes are conversions of data, such as text data, into a form more useful for data analysis.

> For example, a code is usually added to indicate the "outcome" of each case. If there were multiple follow-up phases, the code may indicate in which phase the result was collected. Codes may also be added to indicate editing and missing data actions taken. Text entries are often coded to facilitate analysis. So, a text entry asking for a free form entry of a person's occupation may be coded with a standard code to facilitate analysis.

- Many coding schemes have been standardized.

    Examples: the North American Industry Classification System (NAICS) codes, the Federal Information Processing Standards (FIPS) for geographic codes (country, state, county, etc.), the Standard Occupation Codes (SOC).

Guidelines

- An editing process should be applied to every data collection and to third-party data to reduce obvious error in the data. A minimum editing process should include range checks, validity checks, checks for duplicate entries, and consistency checks.

    Examples of edits: If a data element has five categories numbered from 1 to 5, an answer of 8 should be edited to delete the 8 and flag it as a missing data value. Range checks should be applied to numerical values (e.g., income should not be negative). Rules should be created to deal with inconsistency (e.g., if dates are given for a train accident and the accident date is before the departure date, the rule would say how to deal with it). Data records should be examined for obvious duplicates.

- Most editing decisions should be made in advance and automated. Reliance on manual intervention in editing should be minimized, since it may introduce human error.

- Do not use outlier edits to the extent that special effects and trends would be hidden. Outliers can be very informative for analysis. Over-editing can lead to severe biases resulting from fitting data to implicit models imposed by the edits.

    Rapid industry changes could be missed if an agency follows an overly restrictive editing regimen that rejects large changes.

- Some method should be used to allow after-the-fact identification of edits. One method is to add a separate field containing an edit code (i.e., a "flag"). Another is to keep "version" files, though this provides less information to the users.

- To avoid quality problems from analyst coding and spelling problems, text information to be used for data analysis should be coded using a standard coding scheme (e.g., NAICS, SOC, and FIPS discussed above). Retain the text information for troubleshooting.

- The editing and coding process should clearly identify missing values on the data file. The method of identifying missing values should be clearly described in the file documentation. Special consideration should be given to files that will be directly manipulated by analysts or users. Blanks or zeros used to indicate missing data have historically caused confusion. Also, using a coding to identify the reason for the missing data will facilitate missing data analysis.

  - The editing and coding process and editing statistics should be documented and clearly posted with the data, or with disseminated output from the data.

References

- Little, R. and P. Smith (1987) "Editing and Imputation for Quantitative Survey Data," *Journal if the American Statistical Association*, Vol. 82, No. 397, pp. 58-68.

4.2 Handling Missing Data

Principles

- Untreated, missing data can introduce serious error into estimates. Frequently, there is a correlation between the characteristics of those missing and variables to be estimated, resulting in biased estimates. For this reason, it is often best to employ adjustments and imputation to mitigate this damage.

- Without weight adjustments or imputation, calculation of totals are underestimated. Essentially, zeroes are implicitly imputed for the missing items.

- One method used to deal with unit-level missing data is weighting adjustments. All cases, including the missing cases, are put into classes using variables known for both types. Within the classes, the weights for the missing cases are evenly distributed among the non-missing cases.

- "Imputation" is a process that substitutes values for missing or inconsistent reported data. Such substitutions may be strongly implied by known information or derived as statistical estimates.

- If imputation is employed and flagged, users can either use the imputed values or deal with the missing data themselves.

- The impact of missing data for a given estimate is a combination of how much is missing (often known via the missing data rates) and how much the missing differ from the sources that provided data in relation to the estimate (usually unknown).

> For example, given a survey of airline pilots that asks about near-misses they are involved in and whether they reported them, it is known how many of the sampled pilots did not respond. You will not know if the ones who did respond had a lower number of near-misses than the ones who did not.

- For samples with unequal probabilities, weighted missing data rates give a better indication of impact of missing data across the population than do unweighted rates.

Guidelines

- Unit nonresponse should normally be adjusted by a weighting adjustment as described above, or if no adjustment is made, inform data users about the missing values.

- Imputing for missing item-level data (see definition above) should be considered to mitigate bias. A missing data expert should make or review decisions about imputation. If imputation is used, a separate field containing a code (i.e., a flag) should be added to the imputed data file indicating which variables have been imputed and by what method.

- All methods of imputation or weight adjustments should be fully documented.

- The missing data effect should be analyzed. For periodic data collections, it should be analyzed after each collection. For continuous collections, it should analyzed at least annually. As a minimum, the analysis should include missing data rates at the unit and item levels and analysis of the characteristics of the reporters and the non-reporters to see how they differ. For some

reporting collections, such as with incidents, missing data rates may not be known. For such cases, estimates or just text information on what is known should be provided.

- For sample designs using unequal probabilities (e.g., stratified designs with optimal allocation), weighted missing data rates should be reported along with unweighted missing data rates.

References

- Chapter 4, *Statistical Policy Working Paper 31, Measuring and Reporting Sources of Error in Surveys,* Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget, July 2001.
- The American Association for Public Opinion Research. 2000. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Ann Arbor, Michigan: AAPOR.

4.3 Production of Estimates and Projections

Principles

- "Derived" data items are additional case-level data that are either directly calculated from other data collected (e.g., # of days from two dates), added from a separate data source (e.g., the weather on a given date), or some combination of the two (e.g., give the departing and arriving airports, calculating distance from an external source). Deriving data is a way to enhance the data set without increasing respondent burden or significantly raising costs.

- An "estimate" is an approximation of some characteristic of the target group, like the average age, constructed from the data.

- A "projection" is a prediction of an outcome from the target group, usually in the future.

    Examples: The average daily traffic volume at a given point of the Garden State Parkway in New Jersey two years from now. Total airline operations ten years from now.

- Estimates from samples should be calculated taking the sample design into account. The most common way this is done is weighted averages using weights based on the design.

- Estimates of standard error of an estimate will give an indication of the precision of the estimate. However, it will not include a measure of bias that may be introduced by problems in collection or design.

Guidelines

- Use derived data to enhance the data set without additional burden on data suppliers.

    For example, the data collection can note the departure and arrival airports, and the distance of the flight can be added derived from a separate table.

- Weights should be used in all estimates from samples. Weights give the number of cases in the target group that each case represents, and are calculated as the inverse of the sampling probability. If using weights, adjust weights for nonresponse as discussed in section 4.2.

    > For example, the National Household Travel Survey is designed to be a sample representing the households of the United States, so the total of the weights for all sample households should equal the number of households in the United States. Due to sampling variability, it won't. Since we have a very good count of households in the United States from the 2000 Census, we can do a ratio adjustment of all weights to make them total to that count.

- Construct estimation methods using published techniques or your own documented derivations appropriate for the characteristic being estimated. Forecasting experts should be consulted when determining projections.

    > Example: You have partial year data and you want to estimate whole year data. A simple method is to use past partial year to whole year ratios (if stable year to year) to construct an extrapolation projection (Armstrong 2001).

- Standard error estimates should accompany any estimates from samples. Standard errors should be calculated taking the sample design in account. For more complex sample designs, use replicated methods (e.g., jackknife, successive differences) incorporating the sample weights. Consult with a variance estimation expert.

- Ensure that any statistical software used in constructing estimates and their standard errors use methods that take into account the design of the data collection.

- The methods used for estimations and projections should be documented and clearly posted with the resulting data.

References

- Armstrong, J.S. (2001). "Extrapolation of Time Series and Cross-Sectional Data," in *Principles of Forecasting: A Handbook for Researchers and Practitioners*, edited by J. S. Armstrong, Boston: Kluwer.
- Cochran, William G.(1977), *Sampling Techniques* (3rd Ed.). New York: Wiley.
- Wolter, K.M. (1985). *Introduction to Variance Estimatio*n. New York: Springer-Verlag.

4.4 Data Analysis and Interpretation

Principles

- Careful planning of complex analyses needs to involve concerned parties. Data analysis starts with questions that need to be answered. Analyses should be designed to focus on answering the key questions rather than showing all data results from a collection.

- Analysis methods are designed around probability theory allowing the analyst to separate indications of information from uncertainty.

- For analysis of data collected using complex sample designs, such as surveys, the design must be taken into account when determining data analysis methods (e.g., use weights, replication for variances).

- Estimates from 100% data collections do not have sampling error, though they are usually measuring a random phenomenon (e.g., highway fatalities), and therefore have a non-zero variance.

- Data collected at sequential points in time often require analysis with time series methods to account for inter-correlation of the sequential points. Similarly, data collected from contiguous geographical areas require spatial data analysis.

  > Note: Methods like linear regression assume independence of the data points, which may make them invalid in time and geographical cases. The biggest impact is in variance estimation and testing.

- Interpretation should take into account the stability of the process being analyzed. If the analysis interprets something about a process, but the process has been altered significantly since the data collection, the analysis results may have limited usefulness in decision making.

- The "robustness" of analytical methods is their sensitivity to assumption violation. Robustness is a critical factor in planning and interpreting an analysis.

Guidelines

- The planning of data analysis should begin with identifying the questions that need to be answered. For all but simplistic analyses, a project plan should be developed. Subject matter experts should review the plan to ensure that the analysis is relevant to the questions that need answering. Data analysis experts should review the plan (even if written by one) to ensure proper methods are used. Even "exploratory analyses" should be planned.

- All statistical methods used should be justifiable by statistical derivation or reference to statistical literature. The analysis process should be accompanied by a diagnostic evaluation of the analysis assumptions. The analysis should also include an examination of the probability that statistical assumptions will be violated to various degrees, and the effect such violations would have on the conclusions. All methods, derivations or references, assumption diagnostics, and the robustness checks should be documented in the plan and the final report.

> Choices of data analysis methods include descriptive statistics for each variable, a wide range of graphical methods, comparison tests, multiple linear regression, logistic regression, analysis of variance, nonparametric methods, nonlinear models, Bayesian methods, control charts, data mining, cluster analysis, and factor analysis (this list is not meant to be exhaustive and should not be taken as such).

- Any analysis of data collected using a complex sample design should incorporate the sample design into the methods via weights and changes to variance estimation (e.g., replication).

- Data analysis for the relationship between two or more variables should include other related variables to assist in the interpretation. For example, an analysis may find a relationship between race and travel habits. That analysis should probably include income, education, and other variables that vary with race. Missing important variables can lead to bias. A subject matter expert should choose the related variables.

- Results of the analysis should be documented and either included with any report that uses the results or posted with it. It should be written to focus on the questions that are answered, identify the methods used (along with the accompanying assumptions) with derivation or reference, and include limitations of the analysis. The analysis report should always contain a statement of the limitations including coverage and response limitations (e.g., not all private transit operators are included in the National Transit Database; any analysis should take this into account). The wording of the results of the analysis should reflect the fact that statistically significant results are only an indication that the null

Update
May 2003

hypothesis may not hold true. It is not absolute proof. Similarly, when a test does not show significance, it does not mean that the null hypothesis is true, it only means that there was insufficient evidence to reject it.

- Results from analysis of 100 percent data typically should not include tests or confidence intervals that are based on a sampling concept. Any test or confidence interval should use a measure of the variability of the underlying random phenomenon.

> For example, the standard error of the time series can be used to measure the variance of the underlying random phenomenon with 100 percent data over time. It can also be used to measure sampling error and underlying variance when the sample is not 100 percent.

- The interpretation of the analysis results should comment on the stability of the process analyzed.

> For example, if an analysis were performed on two years of airport security data prior to the creation of the Transportation Security Agency and the new screening workforce, the interpretation of the results relative to the new processes would be questionable.

References

- Skinner, C., D. Holt, and T. Smith. 1989. *Analysis of Complex Surveys*. New York, NY: Wiley.
- Tukey, J. 1977. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
- Agresti, A. 1990. *Categorical Data Analysis*. New York, NY: Wiley.

5. Dissemination of Information

"Dissemination means agency initiated or sponsored distribution of information to the public. Dissemination does not include distribution limited to government employees or agency contractors or grantees; intra- or inter-agency use or sharing of government information; and responses to requests for agency records under the Freedom of Information Act, the Privacy Act, the Federal Advisory Committee Act, or other similar law. This definition also does not include distribution limited to correspondence with individuals or persons, press releases, archival records, public filings, subpoenas or adjudicative processes." – OMB Guidelines.

The first key point in disseminating statistical information is the principle of openness relative to all aspects of quality. Pursuant to that principle, along with the statistical information being disseminated, the related supporting documentation must also be available. That documentation can be separately posted documents referenced by the disseminated information or it can be part of the disseminated entity. The second key point in the dissemination is the final reviews before dissemination. These quality reviews are a final assurance that all quality control steps have been taken and that the dissemination package is complete.

5.1   Publications and Disseminated Summaries of Data

Principles

- In publications or summaries, information should be clearly presented to users, and users should be informed about the source(s) of the information presented.

- As far as possible, tables, graphs, and figures should be interpretable as stand-alone products in case they become separated from their original context.

- Methods used to produce data displayed in tables, graphs, and summary data should be available to the reader.

- Statistical interpretations should indicate the amount of uncertainty.

Guidelines

- Documents should be well organized with language that clearly conveys the message intended. Tables, graphs, and figures should be consistent with each other and the text discussing them.

- All tables, graphs, figures that illustrate data, and text that provides data not in accompanying illustrations should include a title. Titles for tables and graphs should be clearly worded and answer three questions: what (data presented), where (geographic area represented), and when (date covered by data).

- All tables, graphs, figures that illustrate data, and text that provides data not in accompanying illustrations should include a source reference. The source reference should contain one or more entries with references to the sources for the information presented. The reference should be sufficiently detailed for a reader to locate the data used. Since databases and documents may be updated, the "as of" date for the source should also be noted.

- Footnotes should be used, if necessary, to clarify data illustrations, tables, graphs, and figures to clarify particular points, abbreviation symbols, and general notes.

- The style of a publication should conform to specific agency style guidelines to ensure consistency and clarity throughout the document.

- Documents disseminated on the Internet must be accessible as required by section 508 of the Rehabilitation Act (29 USC 794d).

- A contact point should be provided in the publication or with the summaries to facilitate user comments and suggestions.

- Documents containing estimates, projections, and analyses should contain or reference the methodology supporting documentation required in sections 4.3 and 4.4.

  References

- U.S. Government Printing Office Style Manual

5.2   Micro Data Releases

Principles

- The term "micro data" refers to data files with various information at the "unit" level. The unit is dependent upon what data are being collected and from what sources.

  > Examples: micro data may be a collection of individual responses from each person or each household to a survey, reports of information from each company, or reports of individual incidents.

- Making micro data available can enhance the usefulness of the information, and can assist the public in determining whether results are reproducible.

However, micro data should not be released in violation of existing protections of privacy, proprietary information, or confidentiality.

- Micro data should be provided in a manner that facilitates its usefulness to users.

- Quality information as recommended herein, file layouts, and information describing the data (i.e., metadata) enhance the usefulness of the micro data.

Guidelines

- Micro data released to the public should accessible by users with generally available software. It should not be restricted to a single application format.

- Micro data should be accompanied by (or have a reference to) the quality-related documentation discussed herein: planning documentation and collection, processing, and analysis methodology.

- Micro data releases should be accompanied by file layouts and information describing the data.

- Micro data should be accompanied by a clear description of revision information related to the file.

- A contact point should be provided with the data to facilitate user comments and suggestions.

References

- International Standardization Organization standard 11179, Specification and Standardization of Data Elements.

5.3 Source and Accuracy Statements

Principles

- Source and Accuracy Statements (S&As) are compilations of data quality information discussed herein. They provide information on where the data came from, how it was collected, and how it was processed. They include information on known strengths and weaknesses of the data.

- S&As should be regularly updated to include changes in methodology and results of any quality assessment studies.

Guidelines

- The S&A for a data source should contain or refer to the current data system objectives and data requirements as discussed in sections 2.1 and 2.2 of these guidelines.

- The S&A for a data source should contain the data source and data collection design as discussed in sections 2.3 and 2.4 of these guidelines.

- The S&A for a data source should contain or refer to the collection operations methodology documentation discussed in sections 3.2 and 3.3.

- The S&A for a data source should contain or refer to the processing documentation discussed in sections 4.1 – 4.4.

- The S&A for a data source should describe major sources of error including coverage of the target population, missing data information, measurement error, and error measures from processing quality assurance.

- The S&A for a data source should contain or reference the revision process for the system and should indicate the source for revision information for the data source.

References

- General Accounting Office, *Performance Plans: Selected Approaches for Verification and Validation of Agency Performance Information*, GAO/GGD-99-139 (July 1999).
- Office of Management and Budget, *Statistical Policy Working Paper 31: Measuring and Reporting Sources of Error in Surveys* (July 2001).

5.4   Pre-Dissemination Reviews

Principles

- Informal and formal reviews of publications, summaries, or micro data will help ensure that a data product meets a minimal level of quality.

- Due to the diverse aspects of quality in a final product, reviews need to be conducted by several people with different backgrounds.

- Reviews of documentation produced through the various stages of data development will enhance the review process.

Guidelines

- A subject matter specialist other than those directly involved in the data collection and analysis should review the plans, methodology documents, and reports prior to dissemination. They should also review publications and summaries resulting from the data for content and consistency.

- Publications should be reviewed by a style and visual information specialist for compliance with style standards.

- A statistician or other data analysis specialist other than those directly involved in the data collection and analysis should review the plans, methodology documents, and reports prior to dissemination for compliance with these guidelines. They should also review publications and summaries resulting from the data for the wording of statistical interpretation.

- Any items to be disseminated via the Internet must be reviewed by a Section 508 compliance specialist for accessibility.

- Any data products that will be disseminated via special software onto the Internet should be tested for accessibility and interpretability prior to dissemination.

- For micro data releases, the release files and the metadata should be reviewed by an information technology specialist for clarity and completeness.

- If an external peer review process is used: (1) peer reviewers should be selected primarily on the basis of necessary technical expertise; (2) peer reviewers should be expected to disclose to DOT prior technical/policy positions they may have taken on the issues at hand and their sources of personal and institutional funding (private or public); and (3) peer reviews be conducted in an open and rigorous manner.

References

- Ott, E., E. Shilling, and D. Neubauer. 2000. *Process Quality Control: Troubleshooting and Interpretation of Data.* New York, NY: McGraw-Hill.

6. Evaluating Information Quality

Once a data system exists, the key to achieving and maintaining a high level of data quality is to regularly assess all aspects of data quality to improve the data collection and processing procedure or correct special problems with the data as they arise. That can be accomplished by regular assessments of the data collected, special studies of aspects of the data and the effectiveness of the collection and processing methods, and running quality control of key processes to control process quality and collect data quality information.

6.1   Data Quality Assessments

Principles

- "Data quality assessments" are data quality audits of data systems and the data collection process.

- Data quality assessments are comprehensive reviews of the data system to note to what degree the system follows these guidelines and to assess sources of error and other potential quality problems in the data.

- The assessments are intended to help the data system owner to improve data quality.

- The assessments will conclude with a report on findings and results.

Guidelines

- Since data users do not have the same access to or exposure to information about the data system that its owners have, the data system owners should perform data quality assessments.

- Data quality assessments should be undertaken periodically to ensure that the quality of the information disseminated meets requirements.

- Data quality assessments should be used as part of a data system redesign effort.

- Data users, including secondary data users, should be consulted to suggest areas to be assessed, and to provide feedback on the usefulness of the data products.

- Assessments should involve at least one member with a knowledge of data quality who is not involved in preparing the data system information for public dissemination.

- Findings and results of a data quality assessment should always be documented.

References

- General Accounting Office, *Performance Plans: Selected Approaches for Verification and Validation of Agency Performance Information*, GAO/GGD-99-139 (July 1999).

6.2 Evaluation Studies

Principles

- Evaluation studies are focused experiments carried out to evaluate some aspect of data quality.

- Many aspects of data quality cannot be assessed by examining end-product data.

- Evaluation studies include re-measurement, independent data collection, user surveys, collection method parallel trials (e.g., incentive tests), census matching, administrative record matching, comparisons to other collections, methodology testing in a cognitive lab, and mode studies.

- "Critical data systems" are systems that either contain data identified as "influential" or provide input to DOT-level performance measures.

Guidelines

- Critical data systems should have a program of evaluation studies to estimate the extent of each aspect of non-sampling error periodically and after a major system redesign.

- Critical data systems should periodically evaluate bias due to missing data, coverage bias, measurement error, and user satisfaction.

- All data systems should conduct an evaluation study when there is evidence that one or more error sources could be compromising key data elements enough to make them fail to meet data requirements.

- All data systems should conduct an evaluation study if analysis of the data reveals a significant problem, but the source is not obvious.

References

- General Accounting Office, *Performance Plans: Selected Approaches for Verification and Validation of Agency Performance Information*, GAO/GGD-99-139 (July 1999).
- Office of Management and Budget, *Statistical Policy Working Paper 31: Measuring and Reporting Sources of Error in Surveys* (July 2001).
- Lessler, J. and W. Kalsbeek. 1992. *Nonsampling Error in Surveys*. New York, NY: Wiley.

6.3 Quality Control Processes

Principles

- Activities in survey collection and processing will add error to the data to some degree. Therefore, each activity needs some form of quality control process to prevent and/or correct error introduced during the activity.

- The more complex or tedious an activity is, the more likely error will be introduced, and therefore, the more elaborate the quality control needs to be.

- A second factor that will determine the level of quality control is the importance of the data being processed.

- Data system activities that need extensive quality control are check-in of paper forms, data entry from paper forms, coding, editing, and imputation.

- Quality control methods include 100% replication, as with key entry of critical data, sample replication (usually used in a stable continuous process), analysis of the data file before and after the activity, and simple reviews.

Guidelines

- Each activity should be examined for its potential to introduce error.

- The extent of quality control for each activity should be based on the potential of the activity to introduce error combined with the importance of the data.

- Data should be collected from the quality control efforts to indicate the effectiveness of the quality control and to help determine whether it should be changed.

- The quality control should be included in the documentation of methods at each stage.

References

- Ott, E., E. Shilling, and D. Neubauer. 2000. Process Quality Control: Troubleshooting and Interpretation of Data. New York, NY: McGraw-Hill.

6.4   Data Error Correction

Principles

- No data system is free of errors.

- Actions taken when evidence of data error comes to light are dependent on the strength of the evidence, the impact that the potential error would have on primary estimates produced by the data system, and the resources required to verify and correct the problem.

Guidelines

- A standard process for dealing with possible errors in the data system should exist and be documented.

- If a disseminated data file is "frozen" for practical reasons (e.g., reproducibility and configuration management) when errors in the data become known, the errors should be documented and accompany the data.

References

- Ott, E., E. Shilling, and D. Neubauer. 2000. Process Quality Control: Troubleshooting and Interpretation of Data. New York, NY: McGraw-Hill.

6. Evaluating Information Quality

Once a data system exists, the key to achieving and maintaining a high level of data quality is to regularly assess all aspects of data quality and improve the data collection and processing system. That can be accomplished by regular assessments of the data collected, special studies of aspects of the data and the effectiveness of the collection and processing processes, and quality control of key processes to both control the quality during operation and to collect data quality information.

6.1 Data Quality Assessments

Principles

- "Data quality assessments" are data quality audits of data systems and the data collection process.

- Data quality assessments are comprehensive reviews of the data system to note to what degree the system follows these guidelines and to assess sources of error and other potential quality problems in the data.

- The assessments are intended to help the data system sponsor to improve data quality.

- The assessments will conclude with recommendations for data quality improvements.

Guidelines

- Since data users do not have the same access to or exposure to information about the data system that its sponsors have, the data system sponsors should make the initial data quality assessment.

- Data quality assessments should be undertaken periodically to ensure that the quality of the information disseminated meets requirements.

- Data quality assessments should be used as part of a data system redesign effort.

- Data users, including secondary data users, should be consulted to suggest areas to be assessed, and to provide feedback on the usefulness of the data products.

- Assessments should involve at least one member with knowledge of data quality who is not involved in preparing the data system information for public dissemination.

- Findings and results of a data quality assessment should always be documented.

References

- General Accounting Office, *Performance Plans: Selected Approaches for Verification and Validation of Agency Performance Information*, GAO/GGD-99-139 (July 1999).

6.2   Evaluation Studies

Principles

- Evaluation studies are focused experiments carried out to evaluate some aspect of data quality.

- Many aspects of data quality cannot be assessed by examining end-product data.

- Evaluation studies include re-measurement, independent data collection, user surveys, collection method parallel trials (e.g., incentive tests), census matching, administrative record matching, comparisons to other collections, methodology testing in a cognitive lab, and mode studies.

- "Critical data systems" are systems that either contain data identified as "influential" or provide input to DOT-level performance measures.

Guidelines

- Critical data systems should have a program of evaluation studies to estimate the extent of each aspect of non-sampling error periodically and after a major system redesign.

- Critical data systems should periodically evaluate bias due to missing data, coverage bias, measurement error, and user satisfaction.

- All data systems should conduct an evaluation study when there is evidence that one or more error sources could be compromising key data elements enough to make them fail to meet data requirements.

- All data systems should conduct an evaluation study if analysis of the data reveals a significant problem, but the source is not obvious.

Appendix A

References

- General Accounting Office, *Performance Plans: Selected Approaches for Verification and Validation of Agency Performance Information*, GAO/GGD-99-139 (July 1999).
- Office of Management and Budget, *Statistical Policy Working Paper 31: Measuring and Reporting Sources of Error in Surveys* (July 2001).
- Lessler, J. and W. Kalsbeek. 1992. *Nonsampling Error in Surveys*. New York, NY: Wiley.

6.3 Quality Control Systems

Principles

- Activities in survey collection and processing will add error to the data to some degree. Therefore, each activity need some form of quality control system to prevent and/or correct error introduced during the activity.

- The more complex or tedious an activity is, the more likely error will be introduced, and therefore, the more elaborate the quality control needs to be.

- A second factor that will determine the level of quality control is the importance of the data being processed.

- Data system activities that need extensive quality control are check-in of paper forms, data entry from paper forms, coding, editing, and imputation.

- Quality control methods include 100% replication, as with key entry of critical data, sample replication (usually used in a stable continuous process), analysis of the data file before and after the activity, and simple reviews.

Guidelines

- Each activity should be examined for its potential to introduce error.

- The extent of quality control for each activity should be based on the potential of the activity to introduce error combined with the importance of the data.

- Data should be collected from the quality control efforts to indicate the effectiveness of the quality control and to help determine whether it should be changed.

- The quality control should be included in the documentation of methods at each stage.

Appendix A

References

- Ott, E., E. Shilling, and D. Neubauer. 2000. Process Quality Control: Troubleshooting and Interpretation of Data. New York, NY: McGraw-Hill.

6.4　Data Error Correction

Principles

- No data system is free of errors.

- Actions taken when evidence of data error comes to light are dependent on the strength of the evidence, the impact that the potential error would have on primary estimates produced by the data system, and the resources required to verify and correct the problem.

Guidelines

- A standard process for dealing with possible errors in the data system should exist and be documented.

- If a disseminated data file is "frozen" for practical reasons (e.g., reproducibility and configuration management) when errors in the data become known, the errors should be documented and accompany the data.

References

- Ott, E., E. Shilling, and D. Neubauer. 2000. Process Quality Control: Troubleshooting and Interpretation of Data. New York, NY: McGraw-Hill.

APPENDIX B

U.S. Department of Transportation
Request to Seek Correction of Information

In keeping with the non-regulatory nature of these guidelines, this guidance (for the content of requests for correction of information) is not intended to be legally binding requirements. However, DOT may be unable to process, in a timely fashion or at all, requests that omit one or more of the requested elements. DOT will attempt to contact and work with requestors to obtain additional information.

If you need further clarification of the questions below, please refer to the <u>text of these guidelines</u>.

SECTION I: REQUEST FOR CORRECTION

Contact Information:

First Name:

Last Name:

Email:

Organization/Company:

Phone:

Fax:

Title:

Street Address:

City:

State/Province:

Postal Code:

Country:

Describe how the information in question affects you (i.e., how an alleged error harms you, and/or how the correction will benefit you):

```
┌─────────────────────────────────┐ ▲
│                                 │ ▒
│                                 │ ▒
│                                 │ ▼
│ ◄                             ► │
└─────────────────────────────────┘
```

───

Publication Information

Clearly identify the report, data set, or other document that contains the information you want the Department to correct.

DOT Agency	`BTS ▼`
Publication/Report Title	`[                    ]`
Date of report:	`[                    ]`
How did you receive the information? If found on a website, please indicate the URL:	`[                    ]`
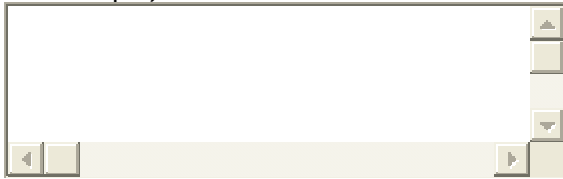
Clearly identify the specific information that you believe needs correction:

```
┌─────────────────────────────────┐ ▲
│                                 │ ▒
│                                 │ ▒
│                                 │ ▼
│ ◄                             ► │
└─────────────────────────────────┘
```

Specify, in detail, why you believe the information fails to meet standards of integrity, utility, and objectivity.

```
┌─────────────────────────────────┐ ▲
│                                 │ ▒
│                                 │ ▒
│                                 │ ▼
│ ◄                             ► │
└─────────────────────────────────┘
```

Specify your recommendations for what corrections DOT should make to the information in question and reasons for believing that these recommended corrections would make the information consistent with both the DOT's and OMB's information quality guidelines.

```
┌─────────────────────────────────┐ ▲
│                                 │ ▒
│                                 │ ▒
│                                 │ ▼
│ ◄                             ► │
└─────────────────────────────────┘
```

In a case where the Department has not designated a report, data set, or document as being subject to these information quality guidelines, and you believe it should be, you should specify why the information should be subject

to the guidelines; and include any documentary evidence you believe is relevant to your request (e.g., comparable data or research results on the same topic).

[text area]
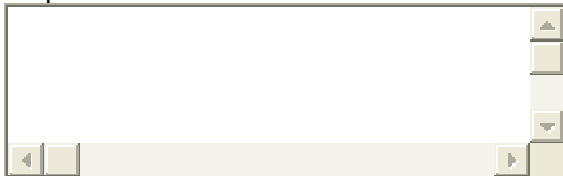
How would you like us to contact you? [Mail ▼]

SECTION II: REQUEST FOR RECONSIDERATION

[ ] Please check this box if this is a request for reconsideration and no more than 30 days has elapsed from the date you received DOT's response to your request for correction. (If this is a request for correction, please complete Section I of this form. You do not need to complete Section II.)

Please provide the following information relating to the request for correction submitted to the Department:

1. DOT Docket Number: OST- [2002 ▼] - [   ]

2. Date request for correction submitted: / [2002 ▼] / [Jan ▼] / [1 ▼]

3. How did you submit request? [Mail ▼]

4. Please provide a detailed explanation of why you are dissatisfied with DOT's response.

[text area]

Privacy Statement | SUBMIT | RESET