DATA VALIDATION AND VERIFICATION STANDARDS

Validation
-Criteria-

□ Goal is appropriate to the identified mission of the organization

- Performance measured has direct bearing (relationship to) on the goal in question

□ Goal is realistic and measurable

- Goal is achievable in the time frame established
- Goal is neither too aggressive in its expectations or set too low for easy achievement

□ Goal is understandable to users

- Terms in goal statement are unambiguous and/or terminology is defined

□ Goal is used in decision making

- Decision makers are identified and their judgment on continued use of the goal in decision-making is periodically evaluated

Validation applies at several levels. First, it is important to establish that the goals that have been selected to measure the performance of the organization have a direct connection and relevance to the mission and desired outcomes that the organization is pursuing. Second, if that relationship is positively established, then the next question to ask is whether the information that is collected clearly relates to the targets that have been set.

Illustration:
If for example, the mission of an organization is to reduce the incidence of a certain disease, it is too indirect to measure the number of brochures that it has distributed to the public about the disease. While this may be educational, it is not a direct indicator of strategy being pursued. It will not inform decision makers of progress in disease eradication, and it may in itself be difficult to measure. For example, while brochures may be distributed to 2,000 centers for distribution, there is no gauge for determining who, if anyone, is taking and reading the information.

Verification
-Criteria-

STANDARDS & PROCEDURES:

□ **Source data are well defined, documented; definitions are available and used**

- Data definitions are well documented and distributed* to all responsible for specific data collection
- Responsible offices can document adherence to data definitions
- Definitions and standards are used in a consistent manner for all parties involved in specific data collection

□ **Collection standards are documented/available/used**

- Protocols and methodology for data collection are documented, distributed* to all responsible for data collection, and adherence to the protocols is required and can be verified
- Data sources are documented

□ **Data reporting schedules are documented/distributed/followed**

- GPRA and other data reporting schedules linked to decision-making are issued* to all parties responsible for data collection; timely data collection and reporting is routinely practiced

□ **Collection staff are skilled/trained in proper procedures**

- Those responsible for either collecting or assembling data are trained for the job. (For data entry responsibilities, see next page)

* distribution can be electronic or hard copy

Illustration:
Standards and procedures refer to establishing the ground rules that should be applied to all data collection efforts for a specific measure. The question is whether the rules are consistently and uniformly applied and clearly communicated to those who are responsible for grass roots data collection. If procedures vary from locale to locale or among individual collectors, results will not be comparable and may not be legitimate. For example, having no clear definition of data to be collected, or mechanisms by which data are collected will inevitably lead to problems in interpreting results or trusting the accuracy of the information. If data definitions are clear, but individuals are not well-trained for the collection effort, which may be complex, additional sources of error may be introduced. Requirements may very well differ from goal to goal, but for a single measure, differences should be minimized if not totally avoided.

Verification
-Criteria-

DATA ENTRY AND TRANSFER

□ **Data entry methodology is documented and followed**

- ° **Documentation of data entry procedures/protocols is available, understood by, and used by data entry personnel**
- ° **Network of data sources is identified**
- ° **Methods used are comparable for all data entry locations**

□ **Data are verified**

- ° **Calculations are checked**
- ° **Data consistency checks are employed e.g. electronic editing**

□ **Procedures for making changes to previously entered data are documented and followed**

□ **Data are available when needed for GPRA reporting and other critical decision making cycles**

□ **Data entry staff are skilled and trained in proper procedures**

Illustration:
Despite the fact that efforts may have been taken to standardize data collection methodology, errors can be introduced when data are entered, transcribed, or transferred during the reporting process. Whether information is being entered into a computer database or being typed up in a report from handwritten notes, errors are possible. The question is whether there is any system in place for detecting these inadvertent errors. Has an office established protocols for checking and approving data that are transcribed in any way? Is there a procedure for addressing the problem of missing data and ensuring that calculations are correct? Does an office employ computer editing systems, when appropriate and feasible, to help identify data entry problems? The use of computer technology to capture data has afforded analytical tools and power that save considerable labor; however, the issue of the accuracy of data being analyzed can be too easily ignored. Identifying data entry or transfer errors is often a very tedious and unrewarding process, but the importance of follow through in this area is nevertheless high

Verification
-Criteria-

DATA SECURITY AND INTEGRITY

□ **Duplicate copies or back-up system for data exists**
- Procedures, including frequency of back up system use, is documented and followed
- Disaster recovery plan in place

□ **Data security protocols are in place and effective**
- Firewalls/password protection, access levels, etc. are established

□ **Equipment and program reliability cannot compromise data accuracy**

Illustration:
This area pertains to precautionary measures that must be taken in the event that computer malfunctions, natural disasters, or human error or actions occur that could affect collected data. Organizations must ensure, whether systems being used are hand-entered records or powerful relational database records, that data are not compromised by lack of attention to security of the data or to the reliability of systems or methods being employed to handle or house data. This means having duplicate records or back up files and ensuring that equipment being employed does the job it was set up or purchased to do. While some problems may be a rare situation, they do occur. For example, some mathematical processes with certain Pentium computers were found to introduce error due to a faulty processor a few years ago. As another example, back-up files, if stored within the same CPU unit on which they were produced, do not offer any additional protection to a malfunction of the hard drive or a fire in that office. These are more indirect considerations for the issue of data accuracy, but cannot be totally disregarded.

Another aspect of data security and integrity is the major question of unauthorized use of data. This could include both external and internal access issues from database “hacking” by external parties, to unauthorized use, including data manipulation, by parties who are not authorized users. A properly designed system will protect internal database systems against unauthorized external use, as well as establish password protection and a clearance process for database changes within the organization.

Verification
-Criteria-

DATA QUALITY AND LIMITATIONS

□ **Accuracy limits of all data are defined**
- Estimated data are identified; methodology for estimation is documented and is supportable; use of estimates are minimized
- Data with margins of error due to accuracy of instrumentation or interpretive leeway, are identified and margin of error (e.g. +/- 1%) is reported.
- Incomplete data are identified and extent of missing data is reported
- Preliminary data are identified and qualifications on data are described

□ **Any other data limitations are explained and documented**

□ **Method for handling anomalous data is established and used**
- Data that appears to be incongruous compared to most other data obtained is re-evaluated and handled appropriately

□ **3rd party evaluations are conducted**
- Objective internal and/or external parties are periodically used to verify accuracy/quality of data
- Use of other crosschecks on data quality such as comparison to similar databases are employed and documented

□ **Use of externally controlled data is minimized**
- Need to use external data is established
- External is identified as such

Illustration:
While every action may be taken to ensure that data are accurately entered, transcribed, reported or otherwise reproduced, there is an underlying question of whether the data itself is accurate or has some inherent limitations. For example, do reports clearly specify that performance may be based on partial data or estimates or that the source of the data is a third party? Do we have any control over such third party data or know how whether the data is accurate or has certain limitations? Are data reported with a confidence interval if that is applicable? Is there a track record for any changes that may have been made over time to the information and why that change was made? In some cases, confidence in data may be bolstered by employing third parties to evaluate the data by peer review, under contract, through an auditing process, or other options. Qualifications on reported data are important pieces of information to decision makers within an agency and in Congress. Recent evaluations of agency performance reports have commended agencies who have explicitly addressed the question of data limitations.

Verification
-Criteria-

OVERSIGHT AND CERTIFICATION

□ **Accountability for data accuracy exists in performance standards**

- Accountability resides with all employees responsible for accuracy of data

□ **Responsible officials certify that procedures were followed each reporting period**

- Signed certifications are filed

□ **Responsible officials certify that data is accurate each reporting period.**

- Signed certifications are filed

Illustration:
The underlying purpose of GPRA is to establish accountability. From the Congressional viewpoint, this means establishing a clear connection between an agency's mission, the work it sets out to do, and what it accomplishes for the funds that have been authorized and appropriated for those purposes. Within each level of an agency, accountability must rest with individuals and officials who are delegated the authority and responsibility for achieving certain goals and striving for specific outcomes. This essentially means that a system of checks and balances are employed to encourage an integrated effort to achieve desired results. Practically speaking, it may often be difficult for management to determine whether information collected or generated by employees is accurate and complete. If they have had extensive prior experience in the area, management may have considerable insight into the processes involved and how to evaluate results. Regardless of the depth of knowledge, staff and management must both be accountable for GPRA data reported. While incorporating accountability into performance standards or employing certifications or attestations to data accuracy will not guarantee that GPRA data is valid and verifiable, such measures will reinforce the importance of accountability and responsibility for performance measurement data and tend to improve the odds that decision makers are dealing with bona fide and reliable information.