

Enforcement and Compliance Data Quality Strategy

March 25, 2002

TABLE OF CONTENTS

	<u>Page</u>
I. Introduction	1
Relationship to Quality Management Plans	1
II. Vision Statement	2
III. Common Data Quality Issues	3
IV. Data Quality Activities	4
A. Periodic Activities	4
Identifying Data Errors	6
Prioritizing Data Problems Found	6
Selecting Data Quality Projects and Involving the Regions and States	7
Sending Analyses of Problems to Regions and States for Correction	7
B. Ongoing Activities	8
Internet Error Correction	8
Promoting Accuracy of Reports, Paper Records, and Data Entry and Validation Procedures	8
Updating Data System Guidance	9
System Updates and ICIS/Modernization	10
V. Regional and State Implementation of the Strategy	10
Coordinating with Other Key Activities	10
VI. Conclusion	11
Appendix 1: The Data Quality Strategy and the Office of Compliance Quality Management Plan	12
Appendix 2: Process for Developing the Enforcement and Compliance Data Quality Strategy and the FY02 Data Quality Strategy Implementation Plan	13
Appendix 3: Overview Chart of the FY02 Data Quality Strategy Implementation Plan	16
Appendix 4: Description and Schedule for Implementing FY02 Projects	17
A. Random Audits	17
B. Comparative Analyses	18
C. Fields Clarification	19
D. Data Quality Alerts	20
Schedule For FY02 DQ Strategy Implementation Projects	21

Appendix 5: Schedule for Implementing FY03 DQ Strategy Projects 23

Appendix 6: Examples of Existing Policies and Procedures for Minimizing Data Errors

- Prior to Data Entry for Our Legacy Data Bases 25
 - A. Region 3 25
 - B. Region 6 26
 - C. Region 7 26
 - D. Region 9 27

I. Introduction

The Enforcement and Compliance Data Quality Strategy's purpose is to develop and implement an ambitious, practical plan to assess and manage the quality of enforcement and compliance data. This work is critical for EPA's own business needs throughout OECA and the Regions, for programs in state and local governments, and for the public. A short but not exhaustive list of reasons the Agency invests in maintaining the integrity of enforcement and compliance data includes the following:

- ▶ High quality data in EPA's national databases is essential for credible measures, reports, and analyses.
- ▶ EPA managers and members of Congress rely on this data to hold EPA programs accountable and to inform their decision-making.
- ▶ Accurate data, in an accessible and usable form, is essential to support Agency planning and targeting.
- ▶ The Government Performance and Results Act (GPRA) requires federal agencies to develop program plans and performance measures, and to report on progress achieved each year. Documenting performance requires timely, high quality data.
- ▶ Accurate and complete information in EPA's enforcement and compliance databases ensures that states are properly credited for their enforcement and compliance monitoring activities.
- ▶ Proof of high quality gives rise to confidence among the public, industry, and other users of enforcement and compliance data, especially if the public is provided Internet access.
- ▶ Conclusions drawn in analyses and reports by OECA or outside groups using EPA data have been challenged based on concerns about data quality. In recent years, Inspector General reports have stressed the need for increased attention to data quality in Agency efforts to monitor and measure enforcement and compliance programs.

Relationship to Quality Management Plans

Although Region's have their own Quality Management Plans (QMP) to ensure data quality, this Enforcement and Compliance Data Quality Strategy covers areas not traditionally covered by Regional QMPs. This fiscal year, OC is also updating its Office-level Quality Management Plan (QMP), which is intended to clearly and fully document the policies, work processes, resources, management structure, and other critical elements of OC's data quality program. For more information on the relationship between this Data Quality Strategy and the OC QMP, please see Appendix 1.

II. Vision Statement

The vision of the Data Quality Strategy (DQS) for enforcement and compliance data is to assure that the data used to support enforcement program decisions are of high quality and accurately reflect the activity and accomplishments of OECA programs. High quality data is defined as accurate, complete, and timely data that are clearly presented and consistent with national data standards. OC believes that assessments of data quality will demonstrate and result in improved accuracy of enforcement and compliance. The Agency can confidently disseminate and use this information to evaluate programs, target resources and to inform Congress, the public, and the news media.

The DQS calls for all levels of OC and its partners to actively manage the quality of data in the national data systems to the highest possible standards by:

- ▶ assessing the quality of enforcement and compliance data in its data systems on an ongoing basis;
- ▶ identifying any missing, incorrect, or inadequate data; and
- ▶ responding to problems as they are identified thereby continually improving data that does not meet data system, program, Office or Agency standards.

The Strategy is designed to reshape *ad hoc* data quality efforts and to establish routine and systematic methods for improving and maintaining a high level of data quality. The focus of this strategy will be to identify data quality problems by examining enforcement and compliance data in Agency systems and raising issues to the appropriate personnel for further analysis and correction. When data quality problems are found to be due to the inherent structure of a database or to established data processing procedures, these issues will be raised both to the appropriate system managers and to the Integrated Compliance Information System (ICIS) modernization team, who can act on many of the recommended improvements in software/hardware specifications, as well as data entry, verification, and documentation.

The following principles will guide implementation of the strategy:

- ▶ **Research** into problems and analyses will focus on real data and its usage.
- ▶ Critical analyses will be **objective**. Analytical techniques that yield comprehensive assessments, such as random sampling of entire data systems, will be used rather than anecdotal investigations. Where practical, root cause analyses will be undertaken to identify systemic reasons for inadequate data quality, which will help formulate practical solutions.

- ▶ **Responsibilities** for finding solutions to data quality problems will be clear. Expectations for performance will be established.
- ▶ Implementation options to improve data quality will be **realistic** and feasible. In some cases, FY02 data quality assessments may examine a longer time period, but implementation options will be limited to remedying only the two most recent years of data.

III. Common Data Quality Issues

The data quality activities discussed in this Strategy are specifically aimed at identifying data that is missing, incorrect, or inadequate. Definitions of these terms are provided below.

Missing Data refer to required files, records or values that are not in the EPA national databases due to incomplete reporting by EPA Headquarters, EPA Region, a state agency, a local authority, or facility. This deficiency precludes accurate program evaluations and could suggest that important environmental work is not being done.

Incorrect Data are values within a field that do not accurately represent the true value. There are three main reasons for data in a required field to be incorrect. These reasons are data creation errors, data entry errors, and ambiguity in use of data fields.

Data Creation Errors: Programmatic staff can err while filling out forms that are forwarded to data entry staff or when entering data online. Some of these errors can be the result of ambiguity in guidance (see below), poorly designed data reporting tools, overlapping systems requiring a staffer to report the same data for entry into two or more systems, and lack of management emphasis on the importance of submitting complete, high quality data. Facilities may also generate errors in preparing required reporting (e.g., monthly discharge monitoring reports required by CWA permits).

Data Entry Errors: Mistakes may be made during data entry by data entry staff or, more rarely, by translator programs. Mistakes may also be due to a lack of data quality checks prior to data entry. For example, established procedures to verify the information on paper records (e.g., inspection reports, enforcement action reports, Case Conclusion Data Sheets) may be inadequate. Finally, there may not be sufficient training and tools provided to staff who enter data to ensure data are entered properly.

Ambiguity in Use of Data Fields: Misinterpretations and/or inadequate guidance concerning what data should be in required fields may result in inconsistency in values across Regions and states and others.

Inadequate Data are defined as records or fields that are not accessible or usable as currently maintained in EPA national databases. For example, data fields might exist in systems, but EPA does not require reporting or has allowed the fields to fall into disuse. Alternatively, new data needs may have arisen due to programmatic changes and the data systems were not modified to meet those needs. In other instances, inadequate data may result if regulations or policies have not been updated to match new goals, technologies, or systems.

IV. Data Quality Activities

This Data Quality Strategy proposes various activities to identify and correct current data quality problems. Annually, OC will develop an enforcement and compliance Data Quality Strategy Implementation Plan in consultation with Regions and states. The goal of the plan will be to identify specific data quality issues through periodic activities, such as audits and analyses. The outcome of these activities will be used to prioritize the order in which to address any data problems. Resulting analyses will be sent to Regions and states for review and correction where appropriate. Ongoing activities, such as the development of online correction, will be undertaken at the same time to address data quality problems that are long-term projects or may require activities on repeating schedules (e.g., training).

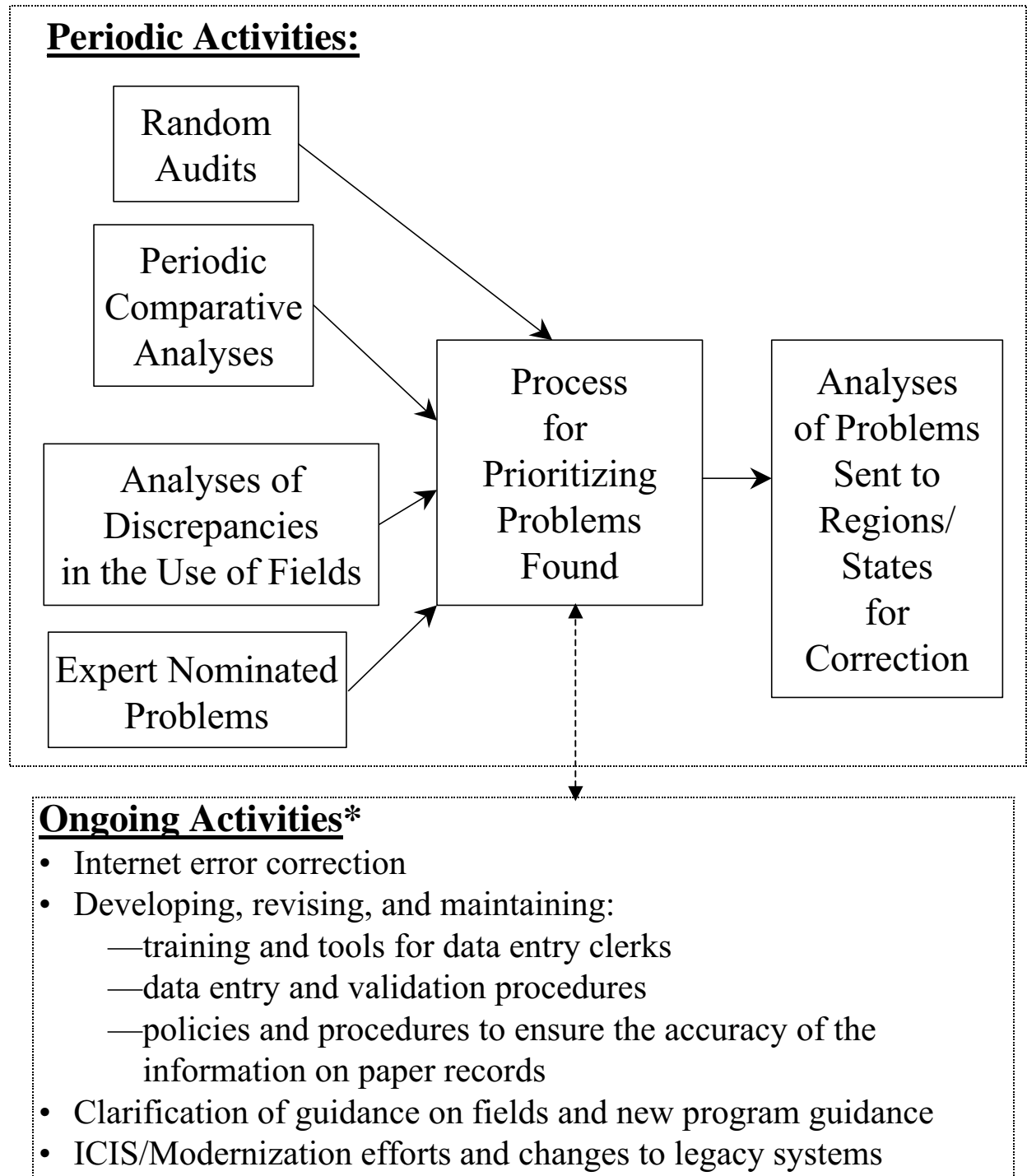
Proposed activities are organized and described in this section according to whether they are periodic or on-going. Figure 2 shows how the activities fit within the DQS and how they are expected to complement each other.

A. Periodic Activities

OC will develop an annual Data Quality Strategy Implementation Plan in consultation with the Regions and states. The plan will identify specific projects for the coming year. It will include relevant details and planning information for:

- ▶ identifying data errors;
- ▶ prioritizing problems found;
- ▶ selecting data quality projects and involving Regions and states; and
- ▶ sending analyses of problems to Regions and states for correction.

Figure 2: Activities Established by the Data Quality Strategy



*Any of these ongoing activities may uncover problems that need to be addressed more formally. In addition, the solutions for some of the prioritized problems may result in revisions or new data entry procedures, clarification of guidance, or system changes/modernization.

Identifying Data Errors

Data problems will be identified through methods such as: random data audits; comparative analyses; analyses of discrepancies in the use of fields; and expert nominated problems. Each of these is described below.

- ▶ Random Data Audits. Data in selected key data fields will be independently evaluated through random data audits to validate its accuracy and completeness.
- ▶ Periodic Comparative Analyses. Periodically, comparative analyses of particular data fields across organizational units with delegated authority will be conducted to identify potential data quality problems. For example, comparing Significant Non-Compliance or inspection rates within Regions or states to the national average can be used to identify extreme outlying values that might indicate data problems.
- ▶ Analyses of Discrepancies in the Use of Fields. OC will analyze key enforcement and compliance activity data fields to determine if there are discrepancies in Regional or state usage. Discrepancies found will be documented, and guidance developed to assist program implementers and database users in how to use the codes for nationally consistent reporting.
- ▶ Expert Nominated Problems. Staff with significant experience and expertise with individual data systems (data system staff, program staff or expert users) will identify any quality issues they encounter. It is expected that such data quality problems will be identified from intensive usage, such as targeting and measures analyses. In addition, data problems found by comparing data required to be entered into more than one database can be used to reconcile information across data systems.

Prioritizing Data Problems Found

Once data problems are identified, they will be prioritized by a standing data quality workgroup that will consult with relevant stakeholders. Data quality efforts will focus on:

- data in the legacy database that will next be modernized;
- administrative priorities of stakeholder agencies and offices involved in management of each data system;
- EPA administration priorities;
- areas of concern raised by the Environmental Council of the States (ECOS), the Inspector General, EPA Management, Regional information users and managers;
- areas of concern discovered by methodologies for identifying data problems (see Figure 2, 'Periodic Activities');
- areas of concern raised by public access; and
- Regional and state review of assessments that confirm validity of the problems.

Selecting Data Quality Projects and Involving the Regions and States

Development of the annual Data Quality Strategy Implementation Plan will follow a structured decision making process. This process will resemble the one used for development of the FY02 Implementation Plan (Appendix 2) and is described below. Each year, OECA's MOA Guidance will specify the number of national data quality projects for EPA and states to work on during each year.

The process will start with an annual Data Quality Strategy workshop held in the spring. Participants will include staff from headquarters, Regions and state representatives. Workshop participants will:

- discuss issues related to the identification and implementation of the current year's projects; and
- brainstorm ideas for the next year's projects.

Sub-workgroups will be convened to build upon the discussions from the Data Quality Strategy workshop and to flesh out the projects to be recommended for implementation as well as methods for analyzing them. During the summer, the sub-workgroups will make their recommendations, which will then be sent to the entire Data Quality Strategy workgroup for review and comment.

The detailed implementation plan for the upcoming year will then be drafted and distributed to a wider set of stakeholders. This implementation strategy will be distributed for review and comment to the entire data steward community, including ECOS, before being finalized. Information on planning, assignment of responsibilities, progress, and results will be coordinated with Regions and states, via the Regions. State and Regions will have scheduled opportunities to communicate their comments on the projects to Headquarters.

Sending Analyses of Problems to Regions and States for Correction

Once the data quality problems are identified via Periodic Comparative Analyses, Analyses of Discrepancies in the Use of Fields, and Expert Nominated Problems and assessments are complete, and problems are researched and well documented, Regions and states will be informed of the problems through distribution of analysis reports in various formats (e.g., short DQ alerts, longer memoranda). These reports will describe in detail an identified data quality problem narratively, and, where applicable, graphically and quantitatively. They will be written for the selected problems and sent to the responsible Regions, states (via the Regions), and Headquarters personnel to alert them of potential data problem so they can correct the data already in the system and make any other changes needed to avoid the problems in the future.

B. Ongoing Activities

The following activities will be undertaken on an ongoing basis to uncover, prioritize and address data quality problems:

- ▶ Internet error correction;
- ▶ promoting accuracy of initial reporting, paper records, data entry, and validation procedures;
- ▶ updating a variety of guidance documents for activities that affect data quality; and
- ▶ updating data system documentation and support system modernization and ICIS development.

Many of these activities will be influenced by and will influence the periodic activities described in the previous section (see Figure 2). For example, data errors identified through periodic audits and analyses may spur changes to ongoing data quality work (e.g., revisions to data entry procedures, clarification of guidance, or system changes/modernization). In turn, the ongoing data quality activities will directly impact the types and number of data quality concerns.

Internet Error Correction

Currently available Internet error correction tools will be applied to particular records in need of correction. For example, OC will use information from the Headquarters, Regional, and state data steward networks, as well as the (Online Tracking Information System) OTIS site (and the Public Access Internet site when it becomes available) and its error correction process to pinpoint data records within OC's national databases in need of correction. The Office of Environmental Information (OEI) is expected to improve the online system for informing data stewards at the Regional and state level about potential problems with individual records reported by data system users. OEI's system will also track the performance of Regional and state data stewards responsible for making corrections. Resources and encouragement from program management will be critical to ensure the continued and effective involvement of those data stewards responsible for responding to error correction requests.

Promoting Accuracy of Reports, Paper Records, and Data Entry and Validation Procedures

Regions and states are expected to commit resources to enter required data into EPA's national databases in a timely and accurate manner. The following are examples of activities OC may undertake, or encourage the Regions and states to undertake, to ensure that data quality procedures are accurately documented and followed.

- ▶ Utilize and document existing edit and validation checks in the legacy databases, and make recommendations for new edit and validation checks to be incorporated into ICIS as the databases are modernized. As part of OC's Quality Management Plan (discussed in the Appendix 1), Data System Quality Assurance Plans will be developed for the each of

the data systems for which OC has primary responsibility. The Data System Quality Assurance Plans will describe the existing and planned editing checks and validation procedures that are part of each data systems.

- ▶ Develop new policies and procedures to minimize data errors stemming from the filling in of field reports and paper records (e.g., inspection reports, enforcement action reports, case conclusion data sheets). Distribute these procedures to the staff responsible for entry. This may include procedures for tracking documents and standardizing document storage. Some examples of existing database policies and procedures for minimizing data errors prior to data entry are listed in Appendix 6.
- ▶ Provide training for all staff generating and entering data that ultimately resides in EPA's data systems. Policies and procedures may need to be reviewed for delegated programs directly entering data into EPA's national databases as well as for states using translator programs.
- ▶ Where the training of data entry personnel is adequate, review the rejected transactions for a specific period of time. In some cases, the data problems may be due to a "process" rather than data entry errors. For instance, duplicate discharge monitoring reports (DMRs) for the NPDES program may be continuously routed for data entry, causing numerous "record on file" rejects.
- ▶ Inform Regional and national program management of the need for sufficient resources to support the implementation of more comprehensive and accurate data entry and verification procedures. Please refer to Appendix 6 for examples of existing policies and procedures for minimizing data errors. For example, disinvestment in accurate, complete data entry and correction for NPDES minors in the Permit Compliance System has eroded the quality of enforcement and compliance data for this universe of over 87,800 facilities.
- ▶ Conduct a one-time review of instructions for data submission to ensure that Regions and states provide data in formats consistent with data system requirements (e.g., values and/or formats should be consistent with data system or translatable with full documentation).

Updating Data System Guidance

Data system guidance and documentation will be updated to address discrepancies in the use of key enforcement and compliance activity database fields. Fields that are inconsistently used by Regions and states could stem from different understandings of how database fields are to be used, or from outdated media-specific program guidance.

System Updates and ICIS/Modernization

OC will work to maintain current, well documented system data dictionaries for those systems run by OC. Such efforts to create and maintain all required documentation will be explicitly called for as part of OC's Quality Management Plan. These data dictionaries should provide clear definitions of data fields and, when necessary, clear definitions of programmatic terms, such as, what a "final judicial order" means for the RCRA program.

OC will also work with the modernization team to ensure that the data quality improvements recommended by the Data Quality Strategy are built into the modernized data systems.

V. Regional and State Implementation of the Strategy

Since Regions and states are responsible for entry of enforcement and compliance information, their active participation in data quality efforts is critical. OECA's Memorandum of Agreement Guidance will continue to specify the Region's level of commitment to implementing the DQS.

OC will propose the following efforts to ensure Regions and states implement the activities of this Data Quality Strategy:

- ▶ Provide yearly quality assurance awards to states.
- ▶ Do not give credit for program actions and results that are not entered in a national database. Notify programs, Regions and state staff and management of this policy.
- ▶ Make any OECA discretionary extramural funding given to states dependent on whether the state commits to and prioritizes maintaining quality data. Maintaining quality data includes providing the data in a form consistent with EPA data system requirements and codes.

Coordinating with Other Key Activities

The proposed data quality planning, implementation, and assessments will be coordinated with other activities to further encourage state and Regional participation in data quality efforts. These activities are listed below.

- ▶ Highlight data quality issues in the Regional profiles that are provided to senior management.
- ▶ Expand program status review reports to include additional data quality information.

- ▶ Publish data quality statistics for Regions to promote positive peer pressure.
- ▶ Communicate findings related to database design issues that affect data quality to the ICIS team.
- ▶ Raise data quality issues through the Enforcement and Compliance Performance Board.
- ▶ Provide annual expectations for data quality goals and/or participation in data quality activities in MOAs.
- ▶ Emphasize data quality in National Environmental Performance Partnerships (NEPPS) and grant guidances.

VI. Conclusion

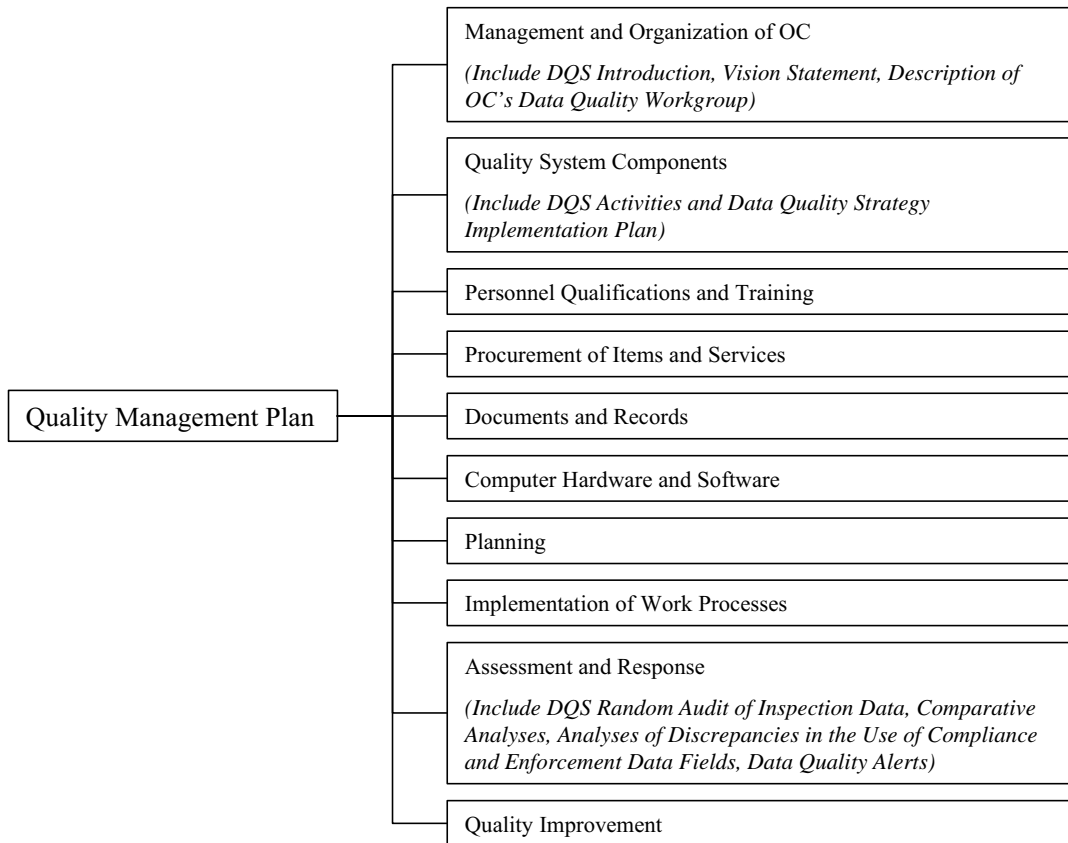
In FY2001, the Office of Compliance began developing this Data Quality Strategy (DQS) for assessing and actively managing the quality of enforcement and compliance information. The DQS sets forth the National Enforcement and Compliance Program's vision for assuring that the data used to support enforcement program decisions are sound and accurately reflect the activities and accomplishments of the program. The DQS also calls for a variety of data quality assessments of EPA data systems which will be undertaken in FY2002 and subsequent years.

Appendix 1

The Data Quality Strategy and the Office of Compliance Quality Management Plan

The updated OC Office-level Quality Management Plan (QMP) will detail how components of the DQS, such as data quality assessments, will inform ongoing planning and implementation of data quality activities and data system management. Figure 1 shows an outline of the major sections in OC's Quality Management Plan and how portions of the Strategy will fit within it.

Figure 1: How the Data Quality Strategy Fits Within OC's Quality Management Plan



Appendix 2

Process for Developing the Enforcement and Compliance Data Quality Strategy and the FY02 Data Quality Strategy Implementation Plan

On May 21, 2001 a call memorandum was sent from Michael Stahl asking for workgroup members to develop a Data Quality Strategy for Enforcement and Compliance Data. Representatives from the media program offices and staff who work directly with EPA maintained enforcement and compliance national databases (DOCKET, AFS, PCS, RCRAInfo, FTTS, and SDWIS) signed up for the workgroup as well as Headquarter personnel (particularly those involved with measures, targeting, and data system management).

On June 20, 2001, a Data Quality Strategy workshop was held in Washington. This served as the first workgroup meeting. There were presentations at the workshop on:

- ▶ an overview document providing a vision statement, definitions, and an initial outline of the Data Quality Strategy,
- ▶ methodologies for identifying, validating, and correcting of data problems and Errors;
- ▶ data errors on the front end; and
- ▶ documentation, guidance, and modernization – their impact on data quality.

In addition, we had brainstorming sessions in the afternoon. Discussions included:

- ▶ various options for conducting the identification and correction of errors;
- ▶ coming up with a list of problems we know about in terms of data entry;
- ▶ coming up with a list of policies and procedures for entering data; and
- ▶ coming of with lists of database fields needing clarification.

On June 28, 2001 summaries of the three breakout sessions were sent to the entire Data Quality Strategy workgroup. On July 6, 2001, the next version of the vision statement, definitions, and outline of the Data Quality Strategy was sent to the workgroup for comment along with documents which grouped the information from the three June 20 workshop breakout sessions

into categories and tasks. We asked workgroup members to get back to us with comments on these documents and the tasks they preferred to work on.

From these responses two sub-workgroups were created:

- 1) Tasks on Methodologies for Identification of Data Problems and Correcting Errors (led by David Sprague)
- 2) Tasks on Potential Fields Needing Clarification: Know Areas where Potential Misuse or Inconsistent Use is Occurring (led by Lynn Vendinello)

We also sent a list of overall, cross-cutting tasks for which we requested input from as many as people as possible.

We did not initiate a sub-workgroup on Data Quality on the Front End. We asked the data system managers to provide us with write-ups describing what data entry and validation checks are already programmed into our national databases. On June 18, 2001, we requested the Data Quality Strategy workgroup members provide us with information on policies and procedures their Regions and states have to ensure the accuracy of the information on the paper records (e.g., inspection reports, enforcement action reports, case conclusion data sheets) **before** they are transmitted to data entry staff. We also requested information on what training and tools are provided to staff who enter data to ensure the data will be entered accurately.

The number of data quality projects which are being recommended for FY02 were decided upon by the data quality and targeting teams of the Information Utilization & Targeting Branch (IUTB), and OC management consensus. The sub-workgroups and the data quality and targeting teams of IUTB recommended projects. Appendix 3 provides an overview of the data quality projects planned for implementation in FY02. Appendix 4 describes the projects in more detail and presents a schedule for implementation. Appendix 5 presents a schedule for FY03 projects. The Potential Fields Needing Clarification sub-workgroup helped decide which fields will be analyzed and how the analyses will be done. The Identification of Data Problems and Correcting Errors sub-workgroup helped develop the Random Audit of Inspection Data project as well as the Comparative Analyses charts. The decisions made by these sub-workgroups were made with consultation from the data quality and targeting teams of the Information Utilization & Targeting Branch (IUTB). In addition, the targeting team of IUTB nominated two data quality alerts to be sent to the user communities early in 2002.

The next step in the review process was to send the recommended data quality projects and analyses to the entire Data Quality Strategy workgroup for review and comment. For example, the entire Data Quality Strategy workgroup were sent:

- 1) the proposals for how to analyze the four codes needing clarification,
- 2) the proposed methodology for conducting the random audit of inspection data,

- 3) sample periodic comparative analyses charts,
- 4) two data quality alerts proposed by the targeting team of IUTB.

Since the Data Quality Strategy workgroup is composed of headquarters and Regional database and program representatives, as well as one state representative, their input was essential.

Once the comments by the Data Quality Strategy workgroup were received, the FY02 Data Quality Implementation Plan was completed. This plan provides detailed outlines of all the data quality projects which will be conducted in 2002. Before finalizing the plan, the Implementation Plan was sent to all the affected user communities, including ECOS, for their review and input.

Appendix 3

Overview Chart of the FY02 Data Quality Strategy Implementation Plan

DQ Activity	Random Audits	Comparative Analyses	Fields Clarifications	Data Quality Alerts
Level of Activity	One per year	Analyses of 4 indicators, three media projects issued semi-annually	New fields are to be analyzed each year	2 issued annually
Current Proposed Project	Inspection Fields of AFS, PCS, and RCRAInfo	Inspection Coverage, formal enforcement actions, penalties and SNC data in AFS, PCS, and RCRAInfo	PCS Enforcement Action Codes 4 RCRA fields	One Air; One RCRA
Correction/ Identification Method Proposed	Select permits/facility ID's randomly. Post inspection data from national data systems for those ID's on website.	State level values compared against national averages	Investigate current usage of data fields and bring findings back to workgroup	Discoveries currently being brought to workgroup attention.
Regional Involvement	Compare paper records to posted data on website. States/Regions can indicate discrepancies online.	Follow-up with states that show low outlier values (e.g., less than half the national ave.).	Workgroup established to develop responses for each field needing clarification.	Clean up data as indicated
State Involvement	Compare paper records to posted data on WEB site. Indicate corrections	States with low outlier values will have to certify validity of data or explain problem.	Workgroup established to develop responses for each field needing clarification.	Clean up data as indicated

Appendix 4

Description and Schedule for Implementing FY02 Projects

The following Appendix describes the projects that OC plans to implement during FY02. An implementation schedule for FY02 projects is presented at the end of this appendix. For further information on any of these projects, please contact David Sprague (sprague.david@epa.gov).

A. Random Audits

Objective: To quantify the accuracy and completeness of selected fields as maintained in the federal data systems.

Example of the type of statements that will be supported by the audit results: "The inspection information maintained by EPA is 95% accurate. Based on the survey methodology used, we are 95% confident that this error rate is within 3% of the true value."

- ▶ This year's audit will examine inspection data only.
- ▶ The audit will focus separately on populations within each of the three media databases:
 - ✓ Facilities subject to CAA programs as defined by the RECAP universe = 42,075
 - ✓ Facilities with active, major water permits = 6,670
 - ✓ Facilities defined as TSDs and LQGs in RCRAInfo = 23,046
- ▶ An audit of 8 facilities per program and state is designed to minimize respondent burden for the Regions and states and will yield statistically valid results with which to characterize accuracy of inspection information in the national data systems.
- ▶ Reviewing this number of facilities should provide a large enough sample size to have 95% confidence in the results, with an error rate of +/- 3%.

Random Audits: Comments and Responses

The following comments were made by states and Regional staff on earlier drafts of the audit methodology:

- ▶ **Comment:** The universe of facilities sampled should include more facilities from larger states and less from smaller states.
Response: In developing the random audit, OC was sensitive to not overburdening individual states. The sample size of eight is based on the number of permit IDs necessary from the largest states (in terms of federally reportable inspections) to obtain results that allow EPA to be 95 % confident that the results are within a reasonably small confidence interval. The sample size of eight does lead to over sampling from smaller

states. This over sampling of some states should provide a cushion by helping to address the possibility that some states will not fully participate. Statistical weighting of the audit results will adjust for over sampling in some smaller states.

- ▶ **Comment:** States should be able to check the hard copy data against their state/local data system at the same time they check the data on the proposed Data Quality Website.
Response: Others in the workgroup believe that it is not always the same person who would check the hard copy and the state/local data systems. To reduce confusion and workload, OC feels that it is appropriate to keep the focus of the random audit on checking discrepancies in the federal data systems. A follow-up request to check state/local data system will be made only if discrepancies are found.

- ▶ **Comment:** States will need notification regarding this effort during the once a year EPA/state MOA or PPG negotiations.
Response: This is a relatively small effort and is directly related to compliance with the Data Quality Strategy. As such, it should be considered as included in the previously negotiated MOAs. In developing the random audit, EPA was sensitive to not overburdening individual states. In any given state program, checking eight records is not expected to be a significant burden.

B. Comparative Analyses

Objective: To identify potential data quality problems (outliers) by comparing state-level statistics to an established benchmark (i.e., ½ national average).

The following activities are proposed for FY02:

- ▶ Comparative analyses charts for RCRA, CWA, and CAA which show all the states sorted by percentages and will include:
 - ✓ Inspection coverage
 - ✓ Ratio of formal enforcement actions to number of facilities inspected
 - ✓ Ratio of HPV/SNC facilities per facilities in violation

- ▶ The Regions overseeing states that fall below a performance threshold (i.e., one-half the national average) would be sent a memorandum asking them to
 - ✓ discuss the findings of the comparative analyses;
 - ✓ certify data is accurate; and
 - ✓ determine cause for missing data.

- ▶ Comparative Analyses: Comments and Responses

Comments by members of the Workgroup included:

- ▶ **Comment:** For air, comparative charts of enforcement actions/total # facilities and HPVs/total # facilities make more sense than having # of inspections and # violations as the denominators.
Response: Using the number of inspections in the denominator is an appropriate measure of data quality, rather than using the number of facilities. This approach to the data quality assessment of actions takes into account two factors that are expected to be related to the reported number of actions: 1) the number of facilities and 2) the number of inspections (the amount of effort put forth by the program). The same logic applies to the SNC analysis. That is, disregarding the level of effort put forth by the program will change the results and highlight states that do not have many facilities in violation.
- ▶ **Comment:** It would be good to coordinate these analyses with the annual audits Regions conduct of states regarding specific facilities.
Response: The Regions are encouraged, as much as possible, to coordinate the concerns raised by these analyses with their audits of their States.

C. Fields Clarification

Objective: Identify discrepancies in the use of key enforcement and compliance data fields.

- ▶ RCRA Fields Clarification: The workgroup of Regional and Headquarters members have participated in a series of conference calls to identify the appropriate use and definition of informal vs formal action codes, and the applicability of corrective action codes as they pertain to the enforcement program. The group has received input from the RCRA and Corrective Action enforcement program offices indicating the appropriate use of existing codes to address enforcement and/or corrective action remediation. A guidance package is being developed and will include specific scenarios to assist program implementers in determining how to use the codes for nationally consistent reporting. Completion of the draft guidance package is expected by mid-March, 2002.
- ▶ Four RCRAInfo codes were examined:
 - ✓ RCRA New Violation or Roll-Over of Existing Violations,
 - ✓ Appropriate Entry of Formal and Informal Actions,
 - ✓ RCRA 500 vs. 600 Series,
 - ✓ Corrective Action Codes in the Enforcement Module.
- ▶ In addition, a DQ sub-workgroup identified one PCS code for study:
 - ✓ Compliance Schedule Event Codes

Fields Clarification: Schedule

- ▶ Donna Inman is working on the necessary guidance and clean-up related to the 4 RCRA codes studied.
- ▶ The PCS Enforcement Action Clean-up guidance will be distributed in final soon.
- ▶ A workgroup has not begun yet to work on clean-up of the compliance event schedule codes in PCS.

D. Data Quality Alerts

Objective: Address data quality problems identified by data users (e.g. targeting, measures, and Regional analysis)

- ▶ Two Data Quality alerts have been chosen:
 - ✓ HPV facilities being entered into AFS without any underlying violations being entered into the system.
 - ✓ Regions and States failing to close out violations in RCRAinfo
- ▶ Two more data quality alerts will be written and distributed later in the year

Data Quality Alerts: Comments and Responses

- ▶ **Comment:** Some reviewers misunderstand the concept of a Data Quality Alert.
Response: Language will be added to the Data Quality Alerts describing their purpose and what responses are expected from the Regions and States.
- ▶ **Comment:** Manual updating of the HPV flag is no longer necessary. The excerpt from Appendix A of the T&A HPV Workbook does not reflect this.
Response: Clarifying language will be added to the CAA data quality alert to respond to these concerns.
- ▶ **Comment:** Facilities could have come into compliance more than 2 years ago but still have HPV status since not all penalties have been paid yet. This effort may prove to be a lot of work with little return if the data is found to be valid for many facilities.
Response: Clarifying language will be added to the CAA data quality alert to respond to these concerns.

Appendix 4: Schedule for FY02 DQ Strategy Implementation Projects (Page 1 of 2)

◆ = Will happen within specified week
 |—| = Will happen over specified weeks
 - - - = Will happen sometime during specified time period

Week of:	2/18	2/25	3/4	3/11	3/18	3/25	4/1	4/8	4/15	4/22	4/29	5/6	5/13	5/20	5/27	6/3	6/10	6/17	6/24	7/1	7/8	7/15	7/22	7/29	8/5	8/12	8/19	8/26	9/2	9/9	9/16	9/23	9/30				
Data review prior to public access announced					◆																																
Data review for public access site						-----																															
Release of public access site																	-----																				
FY02 Audit instructions sent out						◆																															
State and Regional audit of sampled inspection records						-----																															
Distribute, and Regional/State review of FY02 Audit preliminary results															-----																						
Comments on FY02 Audit prelim results due														◆																							
Distribute FY02 Audit final results															◆																						
Clean-up of RCRA fields	-----																																				

Appendix 4: Schedule for FY02 DQ Strategy Implementation Projects (Page 2 of 2)

Week of:	2/18	2/25	3/4	3/11	3/18	3/25	4/1	4/8	4/15	4/22	4/29	5/6	5/13	5/20	5/27	6/3	6/10	6/17	6/24	7/1	7/8	7/15	7/22	7/29	8/5	8/12	8/19	8/26	9/2	9/9	9/16	9/23	9/30		
Distribute, and Regional/State review of 1st Comparative Analysis & “HPVs in AFS” DQ Alert																																			
Responses due from “HPV DQ Alert”																																			
Responses due from 1st Comparative Analyses																																			
Undertake 2nd FY02 Comparative Analysis																																			
Distribute, and Regional/State review of 2nd FY02 Comparative Analysis																																			
Distribute, and Regional/State review of “Closing out violations in RCRAInfo” DQ Alert																																			
Responses due from “Closing out violations in RCRAInfo” DQ Alert																																			

◆ = Will happen within specified week
 |—| = Will happen over specified weeks
 - - - = Will happen sometime during specified time period

Appendix 5

Schedule for Implementing FY03 DQ Strategy Projects (Page 1 of 2)

◆ = Will happen within specified month
 |—| = Will happen over specified months
 - - - = Will happen sometime during specified time period

Month of:	4/02	5/02	6/02	7/02	8/02	9/02	10/02	11/02	12/02	1/03	2/03
Workshops to identify FY03 Data Quality Projects	◆										
Sub-workgroups meet /plan work for FY 03 projects		-----									
Put language in MOAs regarding Regions implementing/overseeing FY03 DQ projects		◆									
Undertake and distribute for review two Misuse of Fields & Comparative Analyses			-----								
Comments due on Misuse of Fields & Comparative Analyses				◆							
Distribute draft of FY03 Implementation Plan for to DQ workgroup for review				-----							
Comments on FY03 Implementation Plan due							◆				
Incorporate comments on FY03 projects							◆				

Schedule for Implementing FY03 DQ Strategy Projects (Page 2 of 2)

Month of:	4/02	5/02	6/02	7/02	8/02	9/02	10/02	11/02	12/02	1/03	2/03
Initiate workgroups to conduct data fields clean-up and guidance efforts								◆			
Undertake FY03 Audit								-----			
Distribute for review FY03 Audit preliminary results										-----	
Distribute FY03 Audit final results											◆
Undertake and distribute for review FY03 Periodic Comparative Analyses & DQ Alerts								-----			
Distribute FY03 Audit final results									◆		

◆ = Will happen within specified month
 |—| = Will happen over specified months
 - - - = Will happen sometime during specified time period

Appendix 6

Examples of Existing Policies and Procedures for Minimizing Data Errors Prior to Data Entry for Our Legacy Data Bases

A few examples of policies and procedures used by Regions and States to minimize data errors during entry into legacy data systems and to ensure the accuracy of information are presented below.

A. Region 3

PCS

Upon return from performing an inspection, the inspector/technical staff member is responsible for completing a regional **inspection form**. The form was created by the PCS Coordinator to collect inspection data and contains fields for collecting permit type information for those facilities that do not have an NPDES permit and are categorized as unpermitted, general, stormwater, etc., making it necessary to create a record in PCS before the inspection can be entered. The form provides a list of all inspection types, inspectors, facility types, and their associated codes. It's purpose is to provide quick and easy access to the information and required fields to enter inspections into PCS. (The inspection form 3560-3 has not yet been updated to include all of the new inspection type codes).

Once inspections have been entered into PCS, the weekly audit reports are checked to ensure acceptance of the inspection. If an inspection was rejected, the problem is resolved and the inspections are re-entered. A retrieval is also pulled to compare the inspections listed in PCS against the inspections sheets that have been collected for the month. For quick checks, an on line verification is used.

A similar process has been set up to collect **enforcement action data** using a created Enforcement Action (EA) Summary Sheet for PCS which list only those enforcement actions used by Region 3 and it's States. The sheet is comprised of enforcement actions and compliance schedules by name and code, it provides the name of the initiator, the date completed, permit number, action date, issued by, docket number, whether it's permitted or unpermitted, associated NPDES number and name, space to list any EA related compliance scheduled events, reason for the EA, and provides Close Out information. This form has been in existence for the past 15 years. It places all necessary information relating to the EA at your finger tips.

To QA enforcement action data in PCS, active enforcement actions and any associated compliance schedules are pulled and circulated among the technical staff. The staff review and update the EA status (e.g., amendments and close outs). This process is performed quarterly.

The Region has shared these procedures and tools with their States. The States of Maryland, Pennsylvania and West Virginia have similar procedures in place for the QA/QC of their data.

B. Region 6

In the RCRA program, after the State and EPA staff return from their inspections, they write up their inspection reports. All of Region 6's state and regional staff have peer review meetings in their Agencies to discuss potential violations. Besides other inspectors/enforcement officers, the Region also invites a lawyer to attend the meetings. Enforcement actions are written and case conclusion data sheets are completed (for EPA only) and attached to the enforcement action before the record is put into concurrence.

The States data (inspection/violation/enforcement) is usually reviewed by a supervisor before it is given to data entry personnel. Each one of the five States operates differently. The States have been informed for many years that they must have all inspection and enforcement actions in the data system at the time that reports are pulled for their mid-year and end-of-reviews. When discrepancies are discovered on the data reports, State personnel are called to resolve them. At the Regional office, all data information pertaining to inspections, violations and enforcement actions are reviewed in the Hazardous Waste Enforcement Branch by a program person who is the liaison with the Regional RCRAInfo group, located in the Information Management Section.

State data entry clerks receive Regional and National training once a year. In addition, they call Region 6's data personnel if they need any help. The Region pulls an Orphan/QA report for States and EPA data once a month. This report denotes data entries that are not properly linked. Also, the Region's data group has developed other reports that they run regularly to check the Region's data.

C. Region 7

CAA

Region 7 does not have any established policy to ensure accuracy of data as it relates to the Air Program. Data entry into the Air Facility Subsystem (AFS) is paper-driven meaning a form is completed by the inspector or case review officer and given to the data entry person to enter the data into AFS. The region performs a cursory review of the data from time to time, but there are no policies in place to correct any data discrepancies.

Docket

When an administrative case has been filed with the Regional Hearing Clerk(RHC)/Docket Analyst, the attorney assigned to the case, completes a Case Initiation Data Sheet (CIDS) and hands this form to the RHC/Docket Analyst, who then enters this initial information into Docket. Upon the conclusion of the case, the staff attorney completes the Case Conclusion Data Sheet (CCDS) which is then given to an attorney for review. After the review process, the form is given to the RHC/Docket Analyst for her to enter the information into Docket. The data contained in Docket is frequently reviewed by the Deputy Regional Counsel. Also if the Docket Manager at Headquarters notices any data discrepancies, he immediately notifies the RHC/Docket Analyst of any mis-information who in turn corrects any data discrepancies. Monthly conference calls and annual meetings are also held.

RCRA

RCRA Information is entered based on what the compliance officers provide. A mechanism used to ensure data quality is RCRA fact sheets. RCRA fact sheets include lists of data elements that the user is supposed to enter into the system, as well as explanations of how the database is structured and how to assure proper data entry. Also, much of the input into the RCRAInfo database is based on letters, notification/inspection forms and telephone conversation records received. The originator of the information is contacted concerning any questions the data analyst may have. RCRAInfo training has been provided to the states. RCRAInfo conference calls are attended monthly along with annual RCRAInfo training.

D. Region 9

Region 9 has built a Lotus Notes system for reporting enforcement data to the DOCKET Coordinator. By eliminating paper forms, and by programming edit checks, help prompts etc. data quality has been significantly enhanced for the most recent two year period in comparison to historical trends.