United States General Accounting Office

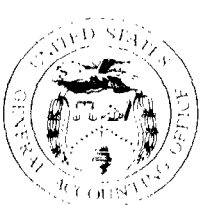
GAO

Report to the Chairman, Committee on Governmental Affairs, U.S. Senate

April 1992

ENVIRONMENTAL ENFORCEMENT

EPA Needs a Better Strategy to Manage Its Cross-Media Information



GAO/IMTEC-92-14

GAO

United States
General Accounting Office
Washington, D.C. 20548

Information Management and
Technology Division

B-246338

April 2, 1992

The Honorable John Glenn
Chairman, Committee on
Governmental Affairs
United States Senate

Dear Mr. Chairman:

EPA's enforcement program is a key means for ensuring that the promise of the nation's environmental laws and regulations is realized. In response to your March 8, 1991, request, this report (1) assesses whether EPA is using sound methodologies to develop information systems that support its enforcement mission by assembling cross-media information from the agency's different environmental programs, and (2) identifies some impediments to EPA's management of information resources to meet users' needs for cross-media information.[1] On June 19, 1991, we testified before your committee on our preliminary observations on these matters.[2] Our objectives, scope, and methodology are detailed in appendix I.

Results in Brief

Deficiencies in EPA's information systems are impeding its ability to enforce environmental laws and regulations. EPA cannot readily bring together and correlate data from its various programs—such as air, water, hazardous waste, and pesticides—in order to assess environmental risks comprehensively or identify and target the most important enforcement priorities.

Two key information systems that were supposed to improve EPA's ability to share and integrate data from among the agency's environmental programs do not satisfy the needs of enforcement users. Compounding this situation are widespread problems with the completeness, accuracy, and timeliness of the data needed by these systems.

These shortcomings are likely to persist unless EPA develops an information management strategy that directly addresses the need for

[1]Cross-media information is derived from data that are pulled together from across EPA's various environmental programs such as those dealing with hazardous waste, pesticides, toxic substances, or air and water pollution.

[2]Ineffective Information Management Impedes EPA's Enforcement Mission and Cross-Media Initiatives (GAO/T-IMTEC-91-16, June 19, 1991).

better integrating the program offices' data and systems. At present, EPA's strategy does not include policies and procedures to help the program offices plan, coordinate, and budget for data sharing. EPA also has not developed a management plan and an information systems architecture that would facilitate such data sharing and integration.

Background

EPA has traditionally enforced environmental laws by identifying violations and taking enforcement actions separately for each environmental medium—air, water, and land—and regulated substance—hazardous waste, pesticides, and toxic substances. Since 1985, however, EPA administrators have articulated a vision of an Environmental Protection Agency managed in a more integrated, comprehensive, and less compartmentalized fashion. The highest priorities of the current Administrator—pollution prevention, management for risk minimization across environmental threats, and a stronger enforcement program—depend explicitly on having accurate, complete information from across the agency and from external sources.

However, in EPA's 1989 and 1990 Federal Managers' Financial Integrity Act (FMFIA) reports, the Administrator stated that the inability to bring together data from across these different media for enforcement purposes constituted a material internal control weakness. According to EPA, this weakness seriously hampered its ability to develop enforcement actions, set enforcement priorities, target enforcement, respond to crises, and conduct program oversight.

The need for improved information systems to support enforcement is also reflected in EPA's October 1990 Enforcement Four-Year Strategic Plan. This plan outlined a fundamental change in the traditional method of media-specific enforcement by emphasizing the need to develop a variety of cross-media data-gathering and analytical capabilities by 1995. Analytical capabilities that are needed include computer systems that support targeting of enforcement activities in specific geographic areas where there are substantial risks to human health or the environment. Geographic areas of concern include high-priority ecosystems such as the Chesapeake Bay and the Great Lakes. Information that EPA needs to support these capabilities and those discussed in the FMFIA reports resides in the data bases of 14 major national information systems. EPA regulates about 500,000 facilities, many of which have records in several of these data bases. However, sharing of information across the agency is difficult because these information systems are largely independent, having been

designed to serve the individual needs of the environmental media program offices.

Prior to 1991, EPA used cumbersome, labor-intensive procedures to perform cross-media analyses of regulated facilities. For example, after the 1989 Exxon Valdez oil spill in Alaska, it took the agency about 3 months to retrieve and manually assemble and analyze the data needed to obtain a cross-media profile of the Exxon Corporation[3] to determine whether there was a corporatewide pattern of noncompliance with environmental laws and regulations. This type of cross-media search has primarily been done by EPA's National Enforcement Investigations Center. Here, skilled analysts compile reports for agency personnel by retrieving data from various agency data bases. Reports generally take from 2 to 4 weeks to produce.

In 1989, EPA's Office of Enforcement (OE) began developing an automated method of obtaining cross-media information, the Integrated Data for Enforcement Analysis (IDEA) system, which is intended to correct the agency's material internal control weakness. While there is no formal plan guiding the development of IDEA, it is essentially being done in two phases. IDEA phase one has been completed and deployed to regional offices for selective access to compliance data in nine major data bases. According to OE officials, EPA spent about $260,000 on phase one. For phase two, which is currently under development, enforcement officials are defining specific analytical requirements, such as those needed to set enforcement priorities or effectively target enforcement. OE intends to develop and deploy these capabilities for use by enforcement personnel throughout the agency, either as part of the IDEA system or a hybrid system that includes an EPA geographic information system. EPA has not developed estimates of the cost or time required to complete IDEA phase two. Funding for continued development of IDEA is contained within a fiscal year 1992 $710,000 budget item called "data integration." Allocations for specific initiatives such as IDEA have not yet been finalized.

IDEA is dependent on another automated system, the Facility Index System (FINDS). FINDS' primary function is to compile lists of the different identification numbers that EPA's major data bases use to refer to a particular regulated facility. EPA's data bases may use different names and numbers to identify the same facility. IDEA then uses the lists of identification numbers to access data on regulated facilities in nine of the

[3]As used in this report, the name "Exxon" includes the Exxon Corporation, its subsidiaries such as the Exxon Shipping Company, and its gas station franchises.

agency's 14 major data bases. Personnel in EPA's National Enforcement Investigations Center also use FINDS as one avenue in their efforts to obtain the identification numbers needed to access data on regulated facilities. FINDS was developed in the early 1980s and is currently being redesigned, in part, to provide lists of identification numbers that are more complete and up to date. EPA's Office of Information Resources Management (OIRM) expects the estimated $1.3 million redesign effort to be completed by the end of fiscal year 1992.

IDEA phase one and FINDS constitute EPA's current cross-media automated information systems capability. In addition, OIRM has started a project called Gateway that will explore options to meet EPA's long-term need for improved access to and analysis of single- and cross-media data. OIRM expects to develop a system as part of the Gateway project that will improve access to its major national data bases and promote data sharing and dissemination throughout the agency.

Systems Development Deficiencies Hamper Cross-Media Enforcement

The methods used to develop IDEA phase one and the redesigned FINDS are not sound or fully consistent with federal guidance or EPA's own systems development policies and procedures. For IDEA phase one, OE deployed the system to EPA's regional offices before adequately testing and documenting it. With regard to the redesigned FINDS, OIRM did not specify users' data quality requirements or adequately plan for future system maintenance. These deficiencies raise serious questions about whether the systems will perform as needed.

IDEA Phase One System Is Inadequately Documented and Tested and Does Not Correct Agency Weakness

OE has not adequately documented the approximately 150,000 lines of instructions in the IDEA phase one software. Consistent with Federal Information Processing Standards (FIPS) guidelines, EPA policy requires systems, such as IDEA, that are intended for wide application to meet critical needs, to be documented thoroughly and formally.[4] Such documentation should include functional and technical requirements analyses; design specifications; diagrams and descriptions of the software logic and data flow; and data storage specifications. Currently, the only documentation OE has for IDEA is a listing of the computer instructions annotated with the programmers' comments. This level of documentation

[4]EPA, OIRM, System Design and Development Guidance, June 1989; EPA, OIRM, Information Resources Management Policy Manual, July 21, 1987; and FIPS Publication 105, Guidelines for Software Documentation Management, June 6, 1984.

is wholly inadequate to ensure that (1) the system is technically sound and (2) someone other than the original programmers can understand, maintain, and enhance the software. The project manager told us he decided to concentrate on developing software rather than properly documenting the system because of resource limitations.

OE officials agreed that the system documentation is inadequate and stated that they plan to better document the software in the coming year. However, the project manager said that he has been unable to find a contractor with sufficient expertise to supplement the three OE programmers developing IDEA and help document IDEA's sophisticated software. Meanwhile, OE is continuing to develop additional analytical capabilities for IDEA before properly documenting the existing software.

In addition, OE deployed IDEA phase one in the agency's regional offices without having a test plan or adequately testing the system. The project manager told us that the only tests of the system were conducted by programmers as they were developing the system, and that the results of these tests were not documented. Agency policy and FIPS publications require that test plans be developed and systems be thoroughly and formally tested before deployment. The test results are to be documented and reviewed to verify that the software operates as intended and ensure that the system effectively and efficiently addresses the agency's information needs.[5] Because the test plan and test results for IDEA were not documented, OE management has no assurance that the system is producing information as intended. Therefore, there is inadequate assurance that IDEA consistently accesses all relevant data in all nine data bases in response to user queries concerning facilities' environmental compliance. The IDEA project manager acknowledged that there was no test plan and that OE did not properly test the system. He stated that, resources permitting, they plan to perform system testing in fiscal year 1992.

In 1991, the Administrator of EPA reported to the President that EPA's material internal control weakness—its inability to associate cross-media data for enforcement purposes—has been corrected as a result of IDEA phase one. However, according to EPA's FMFIA reports, successful correction of this weakness would also include development of an

[5]EPA, OIRM, System Design and Development Guidance June 1989; EPA, OIRM, Information Resources Management Policy Manual, July 21, 1987; FIPS Publication 101, Guideline for Lifecycle Validation, Verification, and Testing of Computer Software, June 6, 1983; and FIPS Publication 132, Guideline for Software Verification and Validation Plans, Nov. 19, 1987.

information system that provides an ability to access data and assess environmental risks, target and prioritize enforcement actions, develop enforcement cases, and plan strategically on a cross-media basis at the state and regional levels. While OE officials told us IDEA phase one provides a basic capability to selectively access cross-media data in 9 of EPA's 14 major data bases and compile that data into one of several output reports, they acknowledged that the system is not being used by regional enforcement personnel for targeting and priority setting. They also agreed that IDEA phase one does not give users the analytical capability needed to routinely set priorities and target enforcement—deficiencies cited in the FMFIA reports. We therefore believe that EPA has not yet demonstrated that IDEA fully corrects its material internal control weakness. We also believe that the IDEA systems development deficiencies discussed in this report substantially increase the risk that IDEA will not be reliable.

In January 1992, enforcement officials told us they believe IDEA has been of significant assistance in developing several new targets for cross-media enforcement actions, including facilities in specific industrial categories, companies with facilities located throughout the United States, and facilities that release specific pollutants. However, they acknowledged that this analysis was done by IDEA's programmers using iterative, experimental procedures, and without specific criteria as to how to rank compliance violations. They agreed that using IDEA for analytical purposes is still too difficult for regional enforcement personnel. IDEA is difficult to use for several reasons: (1) it lacks the ability to rank facilities according to selection criteria, such as number of compliance violations, (2) users cannot transfer selected data elements of interest to other software for further analysis, and (3) users must refer to the nine user manuals for the major national data bases to obtain the data element names and codes needed to retrieve data of interest. A regional enforcement supervisor confirmed that IDEA is still too difficult for the region's enforcement staff to use.

OE officials told us they developed IDEA's data access capability as quickly as possible to meet an urgent need for cross-media data about regulated facilities. They believe they have provided the capability to access nine data bases at a cost of about $260,000 and in a year's time, instead of the considerably greater cost and several years they believe would be required to study, plan, test, and document the system in conformance with federal regulations and guidelines and EPA policies. They said they will assess users' analytical requirements for cross-media enforcement in the coming year, and intend to develop needed analytical capabilities and make the

system easier to use as part of IDEA phase two. However, we remain concerned that IDEA phase one does not represent a sound basis upon which to develop additional analytical capabilities because it has not been adequately documented or tested. Until OE properly documents and tests the system, we believe there is a greatly increased risk that EPA will not be able to maintain or enhance IDEA, and that users will not be able to rely on the system to provide cross-media data.

FINDS Data Quality Requirements Undefined and System Maintenance Unplanned

FINDS' primary function is to provide users with lists of the different identification numbers that the agency's major data bases use to refer to a particular facility. This function is critical because EPA regulates about 500,000 facilities and each data base may use a different name and identification number for the same facility. FINDS links each name and identification number used by different data bases to a unique FINDS identification number and produces a list with this data. FINDS is being redesigned to provide lists of identification numbers that are more complete and up to date, and to help implement the agency's new facility identification policy to standardize identification numbers of each regulated facility in all of the agency's data bases. EPA enforcement officials consider these improvements critical to resolving problems in obtaining accurate and timely information on regulated facilities, since it is impractical for most users to locate all of the enforcement-related data in the agency's numerous data bases without assistance from an automated system.

Within EPA, FINDS is used primarily by analysts at the National Enforcement Investigations Center to obtain facility identification numbers so they can access various agency data bases and compile cross-media compliance reports on facilities and corporations. Users at the Center told us that FINDS data quality problems include the existence of (1) multiple FINDS identification numbers for the same facility, (2) facilities without FINDS identification numbers, and (3) incomplete lists of various names used by the different environmental programs for facilities. These users also told us that because of these data quality problems, they have stopped relying solely on FINDS as a means of identifying and locating information on facilities. In addition, searches across data bases using IDEA reflect the incompleteness of FINDS data. For example, some searches we observed at EPA headquarters showed that there were several hundred facilities without FINDS identification numbers.

While a major emphasis of the redesign effort was to improve FINDS data quality, EPA did not specify users' requirements with respect to data accuracy, completeness, or timeliness. Agency policy requires that users' needs for data quality be considered before developing systems.[6] Specification of users' data quality requirements is especially critical for the redesign of FINDS because of the problems noted above and statements from IDEA system developers that the success of their system is dependent on improved FINDS data quality. Without accurate FINDS identification numbers, IDEA is unable to identify and locate all facilities that may have violation records in the agency's data bases. A senior IRM official told us that during the course of our review OIRM began compiling quarterly reports detailing problems with FINDS data in order to better correct these problems, and OIRM is assessing users' needs for frequency of updates of FINDS data.

EPA also has not adequately budgeted for the maintenance of the redesigned FINDS software. According to FIPS guidance, funding requirements for software maintenance must be anticipated and planned in order to ensure system maintainability.[7] The EPA IRM Policy Manual states that program officials are responsible for planning their budgets to cover the cost of all system components, including the software and data needed to meet users' needs. Since the redesigned FINDS is a critical component of EPA's systems, the agency's budget needs to provide funds to maintain it. The project manager said that a standard provision was made for FINDS maintenance in a general OIRM systems maintenance fund. However, the official acknowledged that maintenance requirements for the redesigned FINDS will be higher than normal because EPA plans to expand the system to cover additional major data bases. A senior IRM official told us that he believes that planned funding will be sufficient to maintain the redesigned FINDS software but will not be sufficient to maintain the facility names and identification numbers. We believe that the lack of a sufficient plan to maintain FINDS increases the risk that funding will not be available to properly maintain the system and users' needs for accurate and timely information will not be met.

[6]EPA, OIRM, Information Resources Management Policy Manual, July 21, 1987.

[7]FIPS Publication 106, Guideline on Software Maintenance, June 15, 1984.

IRM Deficiencies Impede Achievement of EPA's Cross-Media Mission

The EPA Administrator has demonstrated leadership in making the cross-media mission a high-priority management area. However, the agency does not have a complete IRM strategy to achieve its mission. Despite not having this strategy, EPA has begun to plan and develop information systems to support cross-media initiatives such as enforcement. We believe that the systems development deficiencies discussed in this report are due in part to the lack of a complete strategy that would better define requirements for these systems and provide the coordination and resources needed for successful development. We believe that the absence of a complete cross-media IRM strategy greatly increases the risk that efforts to develop longer-range cross-media information systems, such as the Gateway project, will not meet users' needs. EPA's cross-media IRM strategy lacks (1) effective management mechanisms to plan, coordinate, and budget for cross-media IRM activities; (2) an agencywide information systems architecture; and (3) effective cross-media data management.

Ineffective Cross-media IRM Planning, Coordinating, and Budgeting Mechanism

EPA's IRM planning, coordinating, and budgeting have not adequately supported IRM activities, such as managing cross-media data and developing information systems, that are needed to carry out its cross-media initiatives. Such cross-media information management processes are needed to overcome the agency's legacy of independent management of information resources by different environmental program offices. Consistent with Office of Management and Budget Circular A-130, EPA's IRM Policy Manual states that the agency's IRM planning and budgeting must ensure that its acquisition and use of information resources support the mission-based requirements of its program and administrative functions. However, EPA's IRM Policy Manual lacks clear guidance on how to plan, coordinate, and budget for cross-media IRM activities that require support and cooperation from the traditional single-media environmental programs.

Because initiatives such as cross-media enforcement, pollution prevention, and risk reduction depend on data in the various programs' information systems, EPA needs to ensure that its cross-media information requirements are reflected in the plans and budgets of the agency's environmental programs. IRM officials acknowledge that EPA will need to provide additional funding and staff to implement data standards and improve data quality so that the agency's data can be pulled together and analyzed to achieve the agency's cross-media objectives. For example, the manager of the FINDS redesign project stated that the lack of planning,

coordinating, and budgeting for this cross-media system makes it difficult to coordinate and obtain the necessary funding for FINDS data maintenance and quality activities across the agency's regions and environmental programs. The budget director for OIRM also acknowledged that there are no effective cross-media information management processes to provide the IRM planning, coordinating, and budgeting needed to support EPA's cross-media initiatives.

Agencywide Information Systems Architecture Not Developed

EPA has made little progress toward developing an agencywide information systems architecture[8] to provide a standard framework to govern the development, deployment, and use of data and information technology resources in order to accomplish its single- and cross-media missions. This architecture would (1) guide the development of data assets and information systems and facilitate cross-media integration and data sharing among program offices and (2) clearly show how the development and use of information technology, data, and people will support the accomplishment of its cross-media and other environmental protection missions.

The Assistant Administrator for Administration and Resources Management stated that currently there are no plans for developing an agencywide information systems architecture. The Assistant Administrator believes EPA has made substantial progress on acquiring computers and other information technology that would constitute the hardware component of an information systems architecture. However, the agency made these acquisitions without the benefit of an architecture that shows how the acquired technology should fit together to achieve EPA's single- and cross-media missions and goals. Instead, the agency's hardware architecture document simply describes its current environment and several IRM issues that need to be examined, such as IRM training, electronic mail, and image processing. By acquiring hardware before it developed an information systems architecture, EPA has unnecessarily limited its options and increased the risk that its investment in hardware will not meet its needs. Until the agency has an information systems architecture that shows how all of its information resources should fit together to achieve its missions and goals, it will be very difficult to

[8]An information systems architecture is a description of all functional activities to be performed to achieve the desired mission, the elements needed to perform the functions—including all IRM resources (hardware, software, facilities, data, and people)—and the performance levels of those system elements. An architecture includes information on the technologies, interfaces, and location of functions. It is considered an evolving description of an approach to achieving a desired mission.

coordinate the development of systems, such as the system EPA expects to develop as part of the Gateway project, that are intended to provide better access to and sharing of data from across environmental programs.

EPA'S Data Management Not Adequate to Support Cross-media Needs

Although EPA has stressed the need to improve the management of data to support cross-media programs, it has not (1) monitored and improved data quality, (2) implemented data standards, or (3) developed an agencywide data dictionary to meet its cross-media IRM needs. These deficiencies further indicate the need for a complete IRM strategy to overcome the agency's traditional focus on managing data independently for each environmental program.

Poor data quality has been a pervasive problem in EPA for many years. Numerous GAO reports have noted that the lack of accurate, complete, or timely data has adversely affected agency operations (see Related GAO Products). Enforcement officials who use both single- and cross-media data continue to complain that these problems impede their ability to fully identify compliance violations. Senior IRM officials acknowledge that they have not often used their authority or devoted resources to ensure data quality on an agencywide basis. Although senior IRM officials said they intend to pursue improving data quality more aggressively, they have not developed plans to do so. Moreover, they acknowledged that they have not applied sufficient resources to address agencywide data quality problems.

Insufficient development and application of data standards hinders the agency's use of cross-media information and systems such as IDEA, which accesses about 1,000 data elements. For example, EPA has not yet implemented standard definitions for basic data such as the location and identification of regulated facilities, measures of environmental quality, and taxonomy of biological species. Moreover, basic terms such as "enforcement action" and "significant non-compliance" are defined differently by different environmental laws or program offices. Even though the need for agencywide standards was noted by the majority of EPA officials with whom we met, the agency has so far developed just five agencywide data standards—the identification and location of facilities, minimum standards for groundwater data, laboratory data transmission standards, and naming conventions for chemicals—and has not yet fully implemented any of these standards. Senior IRM officials acknowledge that these standards are just the first steps in establishing the needed guidelines to facilitate cross-media integration and sharing of data. Officials said that although their authority to set and enforce data standards is stated in the

agency's IRM Policy Manual, in practice the director of OIRM has little ability to overcome the traditional media-specific method of doing business and ensure compliance on the part of the traditional single-media environmental programs.

The lack of an agencywide data dictionary that contains data names and definitions—such as the data standards discussed above, as well as the multitude of more specialized names and definitions used in the agency's many environmental information systems—makes interpretation of cross-media data difficult. FIPS Publication 76 provides guidelines for developing agencywide data dictionaries so users can readily understand and evaluate different terms, definitions, and sources of data.[9] In order to make effective use of key terms such as "significant non-compliance," users need to know how they are defined by each environmental program. Therefore, in addition to having access to EPA's many data bases, cross-media users must be able to identify and understand data in them. For example, IDEA users currently have access to about 1,000 data elements in nine major data bases. The users must be able to formulate queries using identifying codes and data element names that are unique to each of the data bases. In the worst case, users must refer to manuals that take up about 6 feet of shelf space to formulate queries and interpret results.

Headquarters and regional enforcement officials acknowledged that a data dictionary is needed to permit users to effectively access and use data located in the agency's major national data bases. Senior IRM officials told us that while EPA is developing a data dictionary for administrative operations, there are no specific plans to develop an agencywide dictionary for the environmental programs.

Conclusions

EPA's efforts to bring together data from different environmental programs to accomplish its cross-media enforcement mission and correct its material internal control weakness are jeopardized by systems development deficiencies, insufficient maintenance plans, and inadequate data quality. The cross-media enforcement system does not provide all the capabilities users need to assess environmental risks, target and prioritize enforcement actions, develop enforcement cases, or plan strategically on a cross-media basis at the state and regional levels. As a result, EPA cannot assure that it can identify the most important cross-media enforcement priorities.

[9]FIPS Publication 76, Guideline for Planning and Using a Data Dictionary System, Aug. 20, 1980.

The ability to manage cross-media data is also clearly critical to the success of the agency's cross-media mission, as defined by initiatives such as pollution prevention and management for risk reduction across environmental threats. EPA needs to develop a complete cross-media IRM strategy to effectively develop information systems and manage data to support its cross-media mission. A cross-media IRM strategy will reduce the risk of recurring problems in developing cross-media information systems, such as the problems with IDEA and FINDS discussed in this report, and provide the framework needed for longer-term cross-media systems development efforts, such as Gateway. Until EPA completes its cross-media IRM strategy and develops the information systems needed to support its cross-media mission, the agency will not achieve its vision of protecting human health and the environment through more effective cross-media enforcement and agencywide assessment of risks and prevention of pollution.

Recommendations

To better meet users' cross-media information needs and correct its material internal control weakness by developing IDEA and redesigning FINDS, we recommend that the Administrator of EPA:

- Direct the Assistant Administrator for Enforcement to assess user needs for the IDEA system, including analytical capabilities; develop a formal test plan; properly test existing software; and document the IDEA system design and software before developing additional software for the system.
- Direct the Assistant Administrator for Administration and Resources Management to address data quality problems in the FINDS redesign project by setting standards for accuracy, completeness, and timeliness and by developing a plan for the maintenance of the system.
- Reinstate as a material internal control weakness EPA's inability to bring together data from across the agency for enforcement purposes until EPA demonstrates that its enforcement staff and managers are obtaining the information they need to set enforcement priorities, target enforcement, and conduct program oversight.

To strengthen EPA's overall ability to accomplish its cross-media mission, we recommend that the Administrator of EPA complete the agency's cross-media IRM strategy by

- developing policies and guidance and instituting management procedures to plan, coordinate, and budget for cross-media information resources and activities;

- developing an agencywide information systems architecture that explains the structure of and communications among the agency's information resources that are needed to achieve its single- and cross-media missions; and
- developing an agencywide plan to improve cross-media data quality including setting, implementing, and enforcing data standards and developing and maintaining a comprehensive data dictionary.

As requested, we did not provide a draft of this report to EPA for its review and comment. However, we reviewed the facts contained in this report with the Deputy Administrator of EPA and senior officials in OE and OIRM involved in the issues presented. The Deputy Administrator generally agreed with the facts discussed in our report. He told us that while he believes EPA is making progress in developing a more complete cross-media IRM strategy to support its goals, substantial accomplishments in this area will take time and additional resources. We conducted our review between July 1990 and February 1992 in accordance with generally accepted government auditing standards.

As agreed with your office, unless you publicly announce its contents earlier, we plan no further distribution of this report until 30 days from the date of this letter. At that time we will send copies of this report to the Administrator of EPA, interested congressional committees, the Director of the Office of Management and Budget, and other interested parties. We will also make copies available to others on request. This report was prepared under the direction of JayEtta Z. Hecker, Director, Resources, Community, and Economic Development Information Systems, who can be reached at (202) 336-6416. Other major contributors are listed in appendix II.

Sincerely yours,

Ralph V. Carlone

Ralph V. Carlone
Assistant Comptroller General

Contents

Abbreviations

EPA	Environmental Protection Agency
FINDS	Facility Index System
FIPS	Federal Information Processing Standards
FMFIA	Federal Managers' Financial Integrity Act
GAO	General Accounting Office
IDEA	Integrated Data for Enforcement Analysis
IMTEC	Information Management and Technology Division
IRM	information resources management
OE	EPA's Office of Enforcement
OIRM	EPA's Office of Information Resources Management
PEMD	Program Evaluation and Methodology Division
RCED	Resources, Community, and Economic Development Division

Objectives, Scope, and Methodology

The objectives of our review, as requested by the Chairman, Senate Committee on Governmental Affairs, were to determine (1) whether EPA is using sound methodologies to develop information systems that support its enforcement mission by assembling compliance data from across the agency's different environmental programs and (2) whether there are impediments to the effective management of EPA's information resources to meet users' needs for cross-media information.

To evaluate whether EPA is using sound methodologies to develop systems for cross-media purposes, we (1) examined documents relating to systems planning, design, and development and (2) interviewed OE and OIRM officials responsible for cross-media system development efforts. We evaluated enforcement information systems through (1) examination of documents including strategic enforcement plans, regional enforcement pilot plans, and accomplishments reports and (2) interviews with agency enforcement officials; information management officials; and information system users at EPA headquarters and regional offices, EPA's national office for investigations, and New Jersey's Department of Environmental Protection. We also interviewed enforcement and IRM officials in five EPA regional offices and EPA's National Data Processing Division.

To evaluate impediments to EPA's management of information resources, we reviewed and analyzed agency documents including strategic IRM planning and systems architecture documents, the agency's IRM policy manual, and data policies and standards documents. We reviewed GAO and EPA Inspector General reports that discussed information management and data quality problems at EPA. We interviewed agency IRM officials; the agency's chief statistician; scientists in single- and cross-media program offices at EPA headquarters; and single- and cross-media enforcement officials, systems managers, and users at EPA regional offices.

Interviews and research were conducted at agency headquarters in Washington, D.C.; an agency contractor collocation in McLean, Virginia; regional offices in New York City, New York, and Denver, Colorado; the National Enforcement Investigations Center in Denver, Colorado; and New Jersey's Department of Environmental Protection in Trenton, New Jersey.

Major Contributors to This Report

Information Management and Technology Division

David G. Gill, Assistant Director
Richard D. Eiserman, Assignment Manager
Robert C. Reining, Evaluator-in-Charge
William D. Hadesty, Technical Assistant Director
Prithviraj Mukherji, Technical Assistant Director
Pamela L. Williams, Computer Specialist
Patricia J. Macauley, Staff Evaluator

Related GAO Products

Waste Minimization: EPA Data are Severely Flawed (PEMD-91-21, Aug. 2, 1991).

Toxic Chemicals: EPA's Toxic Release Inventory Is Useful But Can Be Improved (RCED-91-121, June 27, 1991).

Environmental Enforcement: Penalties May Not Recover Economic Benefits Gained by Violators (RCED-91-166, June 17, 1991).

Hazardous Waste: Data Management Problems Delay EPA's Assessment of Minimization Efforts (RCED-91-131, June 13, 1991).

Water Pollution: Greater EPA Leadership Needed to Reduce Nonpoint Source Pollution (RCED-91-10, Oct. 15, 1990).

Disinfectants: Concerns Over the Integrity of EPA's Data Bases (RCED-90-232, Sept. 21, 1990).

Air Pollution: EPA Not Adequately Ensuring Vehicles Comply With Emission Standards (RCED-90-128, July 25, 1990).

Drinking Water: Compliance Problems Undermine EPA Program as New Challenges Emerge (RCED-90-127, June 8, 1990).

Hazardous Waste: EPA's Generation and Management Data Need Further Improvement (RCED-90-3, Feb. 9, 1990).

Superfund: A More Vigorous and Better Managed Enforcement Program Is Needed (RCED-90-22, Dec. 14, 1989).

Air Pollution: National Air Monitoring Network Is Inadequate (RCED-90-15, Nov. 2, 1989).

Water Pollution: Stronger Enforcement Needed to Improve Compliance at Federal Facilities (RCED-89-13, Dec. 27, 1988).

Environmental Protection Agency: Protecting Human Health and the Environment Through Improved Management (RCED-88-101, Aug. 16, 1988).

Ordering Information

The first copy of each GAO report is free. Additional copies are $2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

U.S. General Accounting Office
P.O. Box 6015
Gaithersburg, MD 20877

Orders may also be placed by calling (202) 275-6241.

United States
General Accounting Office
Washington, D.C. 20548

Official Business
Penalty for Private Use $300

First-Class Mail
Postage & Fees Paid
GAO
Permit No. G100