A Report of the

Environmental Information Consortium

For the Office of Environmental Information, US Environmental Protection Agency

June 2005

An Integrated Facility Identification System:

Key to Effective Management of Environmental Information at the Environmental Protection Agency

Facilitated by the

National Academy of Public Administration

Officers of the Academy

Foreword

For the past ten years, the National Academy of Public Administration has conducted numerous independent evaluations of how EPA and state environmental agencies can improve our nation's system for protecting the environment and public health. These studies have identified the need for accurate and comprehensive information as a central element for producing such improvements.

EPA and state agencies need good data to identify environmental problems, set goals, select tools to remedy those problems, and then measure their progress, analyze the effectiveness of programs and adjust management strategies accordingly. Similarly, business managers, environmental groups and researchers all need accurate data to understand facility-specific impacts so they can take appropriate actions to improve the environment and public health.

To advance the development and dissemination of more accurate, timely and consistent information, the Academy agreed to facilitate a broad group of parties interested in improving the collection and management of environmental data. This Environmental Information Consortium (EIC) has included representatives of regulated businesses, environmental and other public interest groups, state environmental agencies, academia and consulting organizations. Together, they have determined that the first priority should be to establish a comprehensive, multi-media approach for integrating facility-specific information across all of EPA's data systems maintained by its media programs.

With EPA funding, the Academy's research staff facilitated the EIC's work. This study has focused on identifying needs that all interested parties have for data that will enable them to understand the multi-media impacts of regulated facilities. In turn, the shared interest in comprehensive data among EIC members has led them to develop recommendations for how EPA can establish an integrated system facility identification system that can form the basis for modernizing EPA's data systems.

EIC members believe this report will be helpful to EPA, Congress, state agencies, the business community, environmental groups and other researchers as they work together to modernize EPA's information systems. The integration of data across all of EPA's media programs is absolutely essential for improving the performance of EPA, state agencies and regulated facilities. Even more important, it is essential for improving how our nation protects public health and environmental quality.

I want to thank EPA for supporting this important work, as well as the EIC's members and Academy researchers, whose cooperation has been so productive and whose diligent work has facilitated the preparation of this valuable study.

C. Morgan Kinghorn
President
National Academy of Public Administration

Table of Contents

Appendices

Acronyms

AIRS	Aerometric Information Retrieval System
AFS	Air Facility System database
CDX	Central Data Exchange
CAFO	Confined Animal Feeding Operation
ECHO	Enforcement Compliance History Online
ECOS	Environmental Council of the States
EDSC	Environmental Data Standards Council
EIC	Environmental Information Consortium
Exchange Network	National Environmental Information Exchange Network
FIS	Facility Identification System
FITS	Facility Identification Template for States
FOIA	Freedom of Information Act
FRS	Facility Registry System
FTE	Full-Time Equivalent
GAO	Government Accountability Office
GIS	Geographic Information Systems
HUC	Hydrologic Unit Code
ID	Identification
IDEA	Integrated Data for Enforcement Analysis
IMWG	State/EPA Information Management Workgroup
IT	Information Technology
NACEPT	National Advisory Council for Environmental Policy and Technology
NAICS	North American Industry Classification System
NEIEN	National Environmental Information Exchange Network
OECA	Office of Enforcement and Compliance Assurance
OEI	Office of Environmental Information
OMB	Office or Management and Budget
PCS	Permit Compliance System
PMA	President's Management Agenda
RCRA	Resource Conservation and Recovery Act
SIC	Standard Industrial Classification
TPA	Trading Partner Agreement
TRI	Toxics Release Inventory

Executive Summary

The **ways** environmental data are reported, collected, recorded, stored, managed, and shared are just as important as **what** data are gathered. While most people agree that collecting and reporting these data are essential for good policy and decision-making, they are now learning that many "on-the-ground" environmental effects and consequences depend on how agencies manage those data and share them among themselves, the regulated community, legislators, citizen groups, and the general public.

Unfortunately, technical constraints and the structure of Environmental Protection Agency's (EPA's) information systems have been sources of complaints almost since Congress first adopted our modern environmental laws. Over the past 30 years, agency data on regulated facilities have been highly fragmented and incomplete. EPA and each state environmental agency—as well as each program office (such as air, water, and waste) within those agencies— separately collect, record, and maintain **both** the specific data needed for each set of regulations **and** the core information to identify each facility by name, address, geographic location, owner, etc.

As a result, regulated facilities have to report and update the same basic identification data to multiple agency programs, at multiple levels of government, and at multiple times over and over again each year or more often. In addition, often there are differences and discrepancies in these identification data that block accurate correlation of the various agency databases.

Yet, a comprehensive and centralized source of facilities' environmental information is essential for the many different parties who are interested in environmental issues. For example, a shared data source can:

- Provide the public a more complete understanding of all the regulatory obligations and environmental impacts at each facility across media programs;

- Save time, staff, and money for both agency regulators and regulated facilities by allowing them to centralize data collection and reporting requirements and to assemble more quickly the multi-media environmental data they need for consolidated reports, permits, and inspections;

- Help agencies to manage their responsibilities more efficiently by identifying the most serious risks, setting priorities, establishing performance measures, and targeting specific companies or facilities for increased monitoring or enforcement;

- Increase data accuracy and thereby reduce the risk of disseminating incorrect information about a company whose name might be improperly linked with facilities that have violated environmental requirements.

In this report, the members of the Environmental Information Consortium (EIC)—who represent the users, collectors, and reporters of these data (see Appendix A)—strongly urge EPA to establish a single shared Master File system that will serve as EPA's official legal record of core identification (ID) data for all facilities regulated under federal environmental statutes.

In 2001, EPA introduced its Facility Registry System (FRS) to reduce the fragmentation of facility data among its regulatory programs. The FRS compiles facility identification records; but it depends on after-the-fact reconciliation of those records, drawn from the separate databases operated by each of EPA's media programs. As a result, FRS records are frequently out of date, incomplete, or inaccurate.

EIC members recommend that EPA should start implementing the Master File system at the national level by requiring all its media programs and regional offices to use and share the core facility ID data in the single Master File system and adhere to uniform procedures necessary to maintain the accuracy of the system. Then, state environmental agencies can participate in the Master File system as they develop the capacity to maintain and update their own facility ID records.

EIC members agree that requiring use of uniform data standards for identifying all federally regulated facilities is essential to operating an effective Master File system. Consistent use of these standards will then allow both EPA's and the states' media programs to share facility ID data that are accurate, complete, and up to date. EPA and the states have already worked together through the Environmental Data Standards Council to adopt uniform standards for facility ID data, and they will continue to update those standards.

The EPA Administrator must require all of EPA's media program databases and all other EPA-funded information systems to use these uniform data standards. In addition, EPA should assist the states with developing their capacity to use the uniform facility ID standards and to share the Master File system, and should eventually require all state agencies to use these uniform data standards for all environmental programs funded by EPA.

For the continued viability and success of the single Master File system, it will also be essential that EPA and the states assure the accuracy of the system's data by working together to adopt clear reporting requirements and data quality objectives. An important feature of these reporting requirements is that facility managers should be able to submit their changes directly to a single Master File. EPA should also evaluate annually the system's data accuracy and publish that evaluation.

Finally, EIC members believe EPA is in the best position to provide the public with access to facility-specific environmental data aggregated across all its media programs and the states. EPA already offers such data through its web site, and EIC members agree that EPA should continue to provide this public service.

The problems and needs identified by the EIC in this report are not new, and there have been many efforts during the past decade that called for a comprehensive facility identification system. But recently, technological advances and voluntary efforts to adopt uniform data

standards have lowered the cost and increased the ability to share data among disparate programs and agencies.

The time is ripe for EPA to make the changes necessary for paving the way toward more effective management of environmental information. With the eight recommendations below, EIC members are calling for EPA to improve how we manage our nation's environmental data in ways that are necessary for achieving more effective and efficient protection of our environment and public health.

Thus, EIC members recommend that EPA take the following actions:

1. EPA should develop, implement, maintain, and oversee a single Master File system for all core facility identification (ID) data and should begin this process as soon as possible.

2. The media program offices at EPA and the state environmental agencies should collaborate in planning, designing, and coordinating how the single Master File system will operate, including: (a) development of uniform procedures and definitions, (b) how the system will track facilities' identities through changes in name, operations or ownership, and (c) how data elements that identify and define each type of facility can support the needs of the various media programs and the state agencies.

3. The EPA Administrator should direct all EPA media programs and regional offices (a) to use and share the single Master File system as EPA's <u>only</u> source of core facility ID data and (b) to assure that future updates or other changes of facility ID data can be directly submitted by facility managers to the Master File system in accordance with procedures jointly designed by EPA and the states.

4. In collaboration with state environmental agencies, EPA should review and modify as needed all the relevant reporting rules or other requirements for its media programs, regions, and state program delegations so the single Master File system is recognized as the official legal record of core ID data for federally regulated facilities; EPA will thereby legitimize use of facility ID data in the Master File system to satisfy all federal data collection and reporting obligations.

5. EPA should require all its media programs and regional offices, as well as state environmental agencies, to adopt and use uniform data elements and standards that identify all regulated facilities, based on the common facility ID elements, definitions, and standards that have been jointly adopted – and will continue to be updated – by the State-EPA Environmental Data Standards Council.

6. EPA should provide assistance to state agencies in adopting and implementing those facility ID standards and in developing the states' capacity to maintain and update their own facility ID records so they can operate effectively in conjunction with the single Master File system.

7.   Every year, EPA should evaluate the accuracy of facility ID data in the single Master File system and then report its findings annually to the states, the public, regulated facilities, and Congress so the system can be continuously improved.

8.   EPA should provide public access to its single Master File system through an on-line, web-based interface that is easy for the public to understand; and that interface should provide for public access to agencies' information about the environmental obligations and performance of all federally regulated facilities.

An Integrated Facility Identification System:
Key to Effective Management of Environmental Information

Introduction

In this report, EIC members call for EPA, in collaboration with state environmental agencies, to establish and share a single integrated Master File system to identify all facilities regulated by federal environmental programs. The need for such a system is not a new problem, nor is the EIC's recommendation new or radical. In 1994, a stakeholder's task force of the National Advisory Council for Environmental Policy and Technology (NACEPT) concluded:

> An integrated information infrastructure with standardized, accurate information that spans the Agency's organizations and its partners is critical to implementation of EPA's guiding principles.[1]

Like the NACEPT Task Force, EIC members come from a variety of organizations that use or report environmental information. They include representatives of environmental and public policy groups, the business community, state environmental agencies, academia, and consulting organizations (see Appendix A). They typify people who submit environmental data to regulators, collect those data, disseminate them, analyze them, or use them in managing or examining environmental impacts.

Despite their differences in outlooks or interests, however, EIC members all recognize the fundamental need for an improved data system to receive, analyze, and update the basic environmental data about regulated facilities that EPA and the states collect. The individual efforts of EIC members over the past several years to promote improved environmental information systems – through their participation in such projects as the Facility Identification Initiative, the Great Printers Project, the State/EPA Information Management Workgroup, and many other studies and initiatives – have provided the background for this report. See Appendix B for a listing of some of the many efforts, past and current, to improve the nation's environmental data systems.

EIC members also share concern and frustration that, despite continued support for improving EPA data systems and many other calls for integrating facility data over the past decade, the necessary administrative, regulatory, and structural changes are not giving appropriate urgency and priority by Congress or the Executive Branch. EIC members once again join their voices calling for change. They hope that this joint effort, despite the members' usually different perspectives, will prompt EPA to cooperate with state environmental agencies and to tackle these information problems aggressively and effectively.

This report focuses on the need to establish and implement an integrated and shared data system that contains core data to identify all federally regulated facilities. Such a facility ID system is the essential first step toward integrating all the environmental information held in the data systems of EPA and state environmental agencies. An effectively integrated and managed

system can accurately identify all federally regulated entities. It can then serve as the foundation for retrieving information across multiple data systems, facilitating public access to that information, and reducing duplicative reporting by businesses.

Early in our research, EIC members learned that adoption and use of uniform data standards and data elements are essential to operating an effective facility ID system, and they are an essential part of data integration, no matter what the structure of the system. While EPA and state agencies have been moving forward on voluntary adoption of uniform data standards, EIC members believe strongly that EPA must mandate their use, and this report includes such a recommendation.

Since 2001, when the EIC first convened, EPA has established a facility identification system called the Facility Registry System (FRS). The FRS is a centrally managed database of regulated facilities that allows users to search for facilities and retrieve selected environmental data contained in the databases of EPA's media programs. While FRS is an improvement over EPA's past data systems, the EIC has found that the FRS, in its current form, falls short of what the public, the regulated community, and environmental agencies need so they can effectively manage and integrate their facility-specific information.

In the following sections, this report explains why our nation's environmental data systems must be integrated. It also describes EPA's current systems and identifies the lessons learned by some states as they develop their own integrated data systems. Finally, this report provides findings and recommendations that, if implemented by EPA, will help the agency to achieve true integration of facility-specific data across multiple environmental data systems.

Section One: Background

A. Punching Holes in the Stovepipes

Environmental laws authorize EPA and the states to collect from regulated entities and other sources a vast amount of information about a broad range of environmental issues. Access to such information is important to protect public health and the environment.[2] Trouble arises, however, in organizing these data and providing statistical reports that are accurate, timely, and easily understood by all. The broad variety of the sources for these data makes it difficult to aggregate and interpret them. In addition, this variety inhibits the uses of adequate data quality controls. These difficulties have long been acknowledged and are well documented.[3]

This data fragmentation reflects the piecemeal evolution of our environmental programs. As Congress enacted environmental statutes to address separate problems – air pollution, water pollution, pesticides, toxic substances, hazardous waste, drinking water, and environmental cleanups – EPA created separate regulatory structures and program offices to implement each new law. Each media office then developed and maintained its own separate information systems and databases for collecting, maintaining, and using media-specific environmental data from regulated entities.

Each media program also created its own numbering system for identifying facilities when they came under regulatory control. With no central coordination, there were different approaches to identifying a regulated entity, separate reporting criteria and formats, and varying reporting cycles. Following EPA's structure, most state environmental agencies have developed separate data systems within the same media-specific "stovepipes."

Congress structured most major environmental programs to be delegated to state agencies, giving EPA broad program oversight and enforcement responsibilities. While EPA is the initial recipient of some data – such as the Toxics Release Inventory, Clean Air Act Risk Management Plans, and pesticide registrations – regulated facilities send the vast majority of their required reports directly to state agencies.[4] In most cases, Congress has required that these agencies provide copies of their data to EPA, but Congress did not mandate standard data formats or reporting schedules.

Over the last decade, three emerging trends have focused a bright light on the fragmented nature of the facility data held by EPA and the states:

- Emerging concepts such as ecosystem protection, pollution prevention, and environmental justice, plus legislative mandates such as the Government Performance and Results Act, have pushed environmental management away from medium-by-medium approaches to more, strategic, multi-media, and performance-based programs.

- Legislative and executive mandates – the Reinventing Government Initiative, Clinger Cohen Act, 2002 E-Government Act, the President's Management Agenda, and the Program Assessment Rating Tool– have demanded greater government efficiency, along

with reducing duplication of effort and burdens for both government agencies and the private sector, through greater use of information technology.

- An Internet-savvy public has increased the demand for greater access to accurate and comprehensive environmental information.

The old stovepipe systems designed to collect and manage environmental data in separate media-specific ways cannot support integrated, comprehensive, and multi-media management of information demanded by these changes.

B. Need for an Integrated Facility Identification System

> *Nine blind men were presented with an elephant, an animal that they knew nothing about. One touched the trunk, another the ears, still another touched the tail, and so on. Afterwards, when asked to describe the elephant, each described only the body part that he had touched because none could see the elephant as a whole animal. Each one knew only part of the truth.*

Much like the limitations of the blind men in this Buddhist parable, we have limited our state and federal environmental data systems to collecting and managing information about a facility through each activity that may produce pollution, rather than the whole entity. As described below, this fragmentation of data presents serious problems and frustrations for regulators, regulated parties, and the public, who all need to report, analyze, and use environmental data.[5]

Federal, State and Local Environmental Agencies

Using our currently fragmented data systems, EPA and state media programs monitor facilities independently to ensure compliance with separate air, water, or hazardous waste requirements. However, these regulators need comprehensive, timely, and accurate information about an entire facility's activities and total environmental impacts. Their needs have gained greater urgency with the emergence of multi-media approaches to environmental protection, such as pollution prevention, performance-based management, multi-media permitting and inspections, and environmental management systems.[6]

With the improved agency-wide analytic capability that could be achieved by integrating facility information, regulators could do their jobs more quickly, efficiently, and accurately. By sharing these data, they could better identify emerging risks, set priorities, and target specific companies or facilities for increased monitoring.[7] Permit writers, enforcement personnel, watershed planners, and others could make smarter decisions about permitting, inspection, and enforcement if they could see a facility's allowable emissions and compliance history across all media programs.

An integrated facility identification system would offer regulators and the public a quicker and more accurate way to compare, for example, the toxic releases reported by companies to EPA's

Toxics Release Inventory with the amounts allowed by state permits. Local emergency planning commissions or state public health offices could more readily identify and track the regulated facilities in their areas and the types of substances used at those facilities that could pose potential risk.

State environmental agencies have already learned that integrated data systems allow them to give the public quicker, more accurate responses when asked for information about specific facilities or specific risks (see Appendix C). These systems have also reduced the time it takes state employees to write permits and prepare for inspections of individual facilities, and produce comprehensive, multi-media analyses of environmental conditions, such as for "state of the environment" reports.

Industry

For federally regulated businesses, the piecemeal approach to reporting and managing environmental information has produced unintended consequences. With each new legislative mandate, the reporting burdens on industry have increased, producing some inconsistent and inadvertent side effects.

Public dissemination of inaccurate data. A vital industry concern is the accuracy of environmental information about their facilities that agencies disseminate to the public.[8] With the final release in August 2003 of EPA's Enforcement and Compliance History Online (ECHO), the public could search for compliance information for over 800,000 facilities nationwide.

During its pilot test of ECHO, EPA found that thousands of errors in facility data on ECHO resulted from inconsistencies in the basic identifying data that was collected from EPA's media-specific databases.[9] Companies have reported ECHO has linked them to violations for which they are not responsible or to facilities they no longer own.[10] The dissemination of incorrect compliance information can have a wide range of adverse effects, from the public's perception of the company to increased insurance rates or questions about qualifying for other business opportunities.[11]

Duplicative reporting requirements. Companies apply for environmental permits to the separate media office in EPA or a state agency, and they often have to report the same information multiple times using different formats. In a poll performed by the National Federal of Independent Businesses in 2003, the third most frequently cited complaint about government paperwork from small businesses was duplicate requests for the same information from various agencies (preceded only by the lack of clear instructions and the volume of paperwork).[12] In an early study of Wisconsin's lithographic printing companies – an industry dominated by small businesses – companies reported that they often spent 150 hours on every major environmental reporting requirement for their operations.[13] Their duplicative reporting requirements among the various media programs were almost solely related to facilities' basic identification data.

Missed opportunities for pollution prevention. Besides creating duplicative reporting requirements, the fragmented approach also has kept some companies from finding opportunities for pollution prevention or addressing regulatory responsibilities more efficiently and at lower

cost. This difficulty is particularly true for small businesses with modest capacity to operate information systems.

The Great Printers Project demonstrated why a unique facility identifier is key to integrating all of a facility's environmental data. In 1996, Environmental Defense, the Council of Great Lakes Governors, and the Printing Industries of America joined together in the Great Printers Project to consolidate printers' media-specific reporting requirements and to link those requirements with information on alternatives for reducing pollution.[14] This project demonstrated that software could be developed that would offer multi-media reporting and permitting requirements to minimize redundancy and confusion for small companies. State agencies reported, however, that a major roadblock to implementing this new approach was the lack of a computer system that would pull together and integrate all the data about a single facility from among the various media-specific databases.

Community, Public Interest, and Environmental Groups

Environmental conditions can put the health of the public and workers at risk, so people need and want accurate, comprehensive environmental data about the places where they work, live, and play. They want to learn more about their local environment, potential exposures to pollution, and about the backgrounds of companies located in their community. They want answers to basic questions, such as what facilities are located in a particular community, city, zip code or watershed, or what types and amounts of pollution they are emitting.

Like industry, the public wants that data to be accurate and timely, and they want it in a format that is easily understood and digested. These needs apply whether the public wants to prepare for a specific facility's permit hearing or renewal, to assess the impact of siting a new facility, to evaluate the environmental impacts of new regulations, or to decide whether to purchase a home in a particular location.[15]

Academic and Government Researchers

Research provides critical information about relationships among actions by regulated entities and health or environmental outcomes. Researchers help to assess which types of agency actions have beneficial effects and which do not. For example, researchers evaluate facility-specific data to determine if there are links between exposures to specific toxic releases and progression of heart disease, or they analyze the consequences for human and ecological health of air- and water-borne pollution from particular facilities. With easy access to computer technologies that facilitate tracking environmental impacts, researchers are pursuing these studies much more frequently and on a nationwide basis.[16]

Researchers, both within EPA and in private institutions or environmental groups, use facility-specific information from multiple EPA databases for their projects, and they have reported that they currently must spend a significant portion of their research budgets on efforts to match facilities and "clean up" duplicate or missing facility data from EPA's media specific systems.[17] Thus, they have expressed the need not only for a unique facility identification system that can

link all facility-specific information, but also for more accurate and timely data to identify regulated facilities in EPA's media programs' databases.[18]

Other Government Agencies

Many other local, state and federal agencies have particular interests in facilities regulated by EPA. Although the specific uses of EPA data for the purposes of other federal agencies is beyond the scope of EIC's research, EPA has developed an awareness of need for such data by local emergency response officials.

For example, local emergency management officials might need to know the inventories of hazardous chemicals held by facilities so they can ensure their security and better prepare for potential accidents. An integrated facility identification system that would make it easier to identify agencies' data about any specific facility would save both regulated facilities and other agencies a large amount of staff time and other resources by keeping these data current and by assuring they can share accurate information in a timely manner.

C. EPA's Efforts to Integrate Facility Data

The Early Years

For many years, EPA has recognized the need for an integrated data system that could pull together and link all of the fragmented pieces of environmental information about regulated facilities contained in all of its media programs' databases. Since the 1980's, EPA has tried several times to develop a centralized facility identification process that would link all of these records (see Appendix B).

EPA tried one computer application that sorted through its major program databases to find similar-looking names and addresses so it could assign a single number to each facility.[19] In theory, this facility number could then provide access to information from all the media-specific databases. In 1991, this system served as the linking mechanism for EPA's first application that enabled multi-media analysis of facility data – the Integrated Data for Enforcement Analysis (IDEA)[20] – developed by the Office of Enforcement and Compliance Assurance (OECA).

IDEA's linking mechanism was helpful in pulling together multi-media data about regulated facilities, but there were many problems and errors with the accuracy of the data it produced.[21] In 2000, a study prepared for EPA's Facility Identification Initiative found that IDEA had particular difficulty integrating data because the underlying facility identification information – such as the name, address or SIC code – in the various programs' databases was often inconsistent.[22] These differences made IDEA's "passive," computerized process for matching facilities' data especially prone to inaccuracies.[23]

The 2000 study recommended that:

- EPA move away from this "back-end" or after-the-fact integration[24] of environmental data toward a "front-end" approach that would allow facilities to report their own basic facility information and multi-media interests to a central database;

- Information in that central system then should be checked for completeness and accuracy of linkages by EPA or state agency staffs;[25]

- Data quality should be improved by having the facilities report their own information to one place that would be the source of their data for all parties; and

- The initial reporting burden on regulated facilities would likely be offset in the long run by reducing the number of forms they would otherwise be required to submit to all the separate media programs in EPA and the states.[26]

Instead of eliminating the back end approach as recommended in that study, however, EPA decided to avoid changing the media programs' reporting and data management practices. Rather, EPA attempted to build a hybrid system, creating a central database of facility records and linking environmental interests or data from the various media programs' databases. But EPA still created and maintained this system solely through after-the-fact (or "back-end") reconciliation of facility data.

EPA's Facility Registry System

EPA's current facility identification system that links the separate databases of its media programs is called the Facility Registry System (FRS). FRS began operation in 2001 as part of EPA's newly created Office of Environmental Information (OEI). FRS's central database currently contains records for more than 1.5 million facilities and incorporates over 2 million separate program identifiers.

FRS works on the same principle as EPA's prior linking mechanism. It is based on back-end reconciliation of facility data that were previously entered the various media programs' data systems. FRS's computer application scans the facility records contained in all of EPA's media-specific databases to look for duplicate records on the same facility. FRS generally uses the physical location of a facility as found in the underlying databases to select and match its records.[27] The media programs' databases are not affected by the FRS system, and the program staffs do not have any need to use or update the FRS's facility records.

For each facility, FRS compiles a record of all identifying information, which may include name, location, affiliated organizations, separate media program identifying numbers, industrial codes, different mailing addresses, contact persons, alternative names used by each of the program databases, and other identification data such as the hydrologic unit code (HUC), tribal lands, and EPA region.[28] FRS then randomly generates a unique FRS identifier number and assigns it to that facility record.

FRS's current matching software is fairly sophisticated. For example, it may be able to determine that, when the air program lists a facility at 100 Main Street and the water program identifies a facility at Main and First Street, they are the same physical location and should be combined as the same facility under one FRS record and identifier number. When addresses are similar, but not identical, such as 100 South Main and 100 Main, the application chooses a default for the FRS record the address listed in a facility's annual Toxics Release Inventory form, if it reports to that database.[29]

To update FRS records, its software periodically "extracts" or scans and retrieves updated information from EPA's program databases on a monthly, quarterly, yearly, or bi-yearly basis, depending on the database and the reporting cycles.[30] FRS automatically flags and picks up any changes in facility data contained in one of the programs' databases and then adds them to its records. However, FRS does not normally alert other program offices with an interest in a facility about any changes in its FRS record.

The current FRS is clearly an improvement over EPA's past efforts to integrate facility-specific ID data, and OEI has worked hard to improve the quality of FRS records. FRS is now a core component of EPA's enterprise architecture – the basic structure of its information systems – that was recently approved by the Office of Management and Budget.[31] Improved technology and additional funding for EPA to oversee and correct facility data have made FRS records more accurate than ever before. Thus, FRS provides a better linking tool than EPA's prior systems, although its back-end reconciliation process is still very labor intensive and rather slow, often making the data out of date.

Other Efforts to Improve Environmental Data

EPA's Office of Environmental Information and state agencies have also been working in other ways to improve environmental data management, integration and sharing of information.

Public Access to Multi-Media Data Searches

FRS makes it possible for the public and agency staff to conduct multi-media data searches and gain access to mapping applications – such as ECHO, Envirofacts, IDEA, and others[32] – through EPA's website. These on-line searches rely on FRS records to retrieve multi-media information from EPA's underlying media-specific databases. For example, through ECHO, FRS can answer one request for a search of all EPA enforcement actions or notices of violations against a particular facility, whether from the air, water, or hazardous waste programs.

Error Notification

EPA's websites also offer a link for on-line notification of errors, which allows users to alert EPA about possible errors in FRS data. EPA's Integrated Error Correction Process, which is maintained by OEI, sends error notices to an appropriate agency staff for investigation.[33] EPA data stewards in headquarters and the regional offices then manage, research, and correct any reported errors in FRS records. Their research sometimes involves using maps, telephone books,

information from inspectors, or direct contacts with facilities to find the correct facility name or address.

Standard Data Elements

Groups such as the State/EPA Information Management Workgroup (IMWG)[34] have exerted their influence and pressure to facilitate sharing and integrating environmental data among state and EPA data systems. Many state agencies are now following at least some of the voluntary guidelines in two versions of the Facility Identification Template (FITS I and II), developed initially in 1997 and updated in 2000.[35] In November 1999, the IMWG established the Environmental Data Standards Council (EDSC) to identify, develop, and promote the adoption of data standards that will improve consistency and quality in exchanging information among partner agencies.[36] The EDSC, which develops the data standards that are used by participants in the National Environmental Information Exchange Network, has approved and adopted twelve final data standards, including one for facility identification.[37]

National Environmental Information Exchange Network

To help in resolving the problem of incompatible data systems and incompatible formats among the states and EPA, OEI and state agencies developed the National Environmental Information Exchange Network (Exchange Network), which became operational in 2003.[38] With the Exchange Network, agencies can send and receive environmental data even if they do not have same types of data systems. Exchange Network partners sign Trading Partner Agreements (TPAs) in which they commit to using the standard data elements developed by the EDSC so they can share data regardless of their computer systems or platforms.

Each participating agency develops its own "node" for the Exchange Network, which is securely connected to the Internet and serves as the exchange point for all data requests and submissions to that agency. EPA's node or central point of contact for all of its different media data systems is called the Central Data Exchange (CDX). The CDX provides the technical capability that allows facilities to submit their reports directly to EPA through the Internet, which they can now do for the Toxics Release Inventory (TRI) reports.

Participation in the Exchange Network is voluntary, although EPA provides grants to encourage states to join. As of June 2005, 33 states have operational servers that allow them to access the Exchange Network; but most states currently use the Exchange Network for only a few of the different data types that flow between the states and EPA.[39] Twenty-four states exchange facility records with EPA's FRS through the Exchange Network. For a few states that are actively managing their own facility ID data, FRS automatically chooses the state data for the FRS record if there are inconsistencies.[40]

Project to Improve Locational Data

In 1996, EPA launched an effort to improve its data about the location of entities with any environmental interest. Using new geographic information system (GIS) tools, EPA and the states have begun a systematic process for cleaning up location data on all regulated facilities.[41]

D. State Facility Identification Systems

In researching this report for EIC members, Academy staff asked a number of state environmental agencies for detailed information about their facility identification systems. Academy staff used e-mails and telephone calls to interview Chief Information Officers (or equivalent positions) in 16 state agencies, most of who were members of the State/EPA Information Management Workgroup. See Appendix C for a Summary of State Responses about State Facility ID Systems.

The information gathered from these interviews does not provide a comprehensive assessment of state data systems. These 16 states do, however, represent a diverse mixture of population, land size, geography, urban and rural economies, and widespread location throughout the country. As a result, the interviews provided helpful insights to EIC members for this report.

Among the 16 states, 12 are currently operating some type of central facility ID system to integrate facility-specific information from the different state media programs. Three more states are now in the process of developing such systems. One state has a computer system that is capable of tracking facility-level information across all of its media programs, but it does not currently use this function.

Most states that have established central ID systems have been driven by the need for agency staff and managers to track all environmental activities at each regulated facility, including data on inspections, compliance, and enforcement. These states also wanted to improve their data quality, reduce data redundancy, and improve overall efficiency in managing facility data shared by multiple programs. One state took advantage of the redevelopment of its IT systems in preparation for Y2K to install a more integrated database of facility ID information.

All 12 states currently operating central facility ID systems use them for uniquely identifying each facility and for linking data about each facility from separate systems maintained by their media programs. These systems all store core records of general ID information about regulated facilities and include links to more media-specific data in underlying state program databases. All of these state systems also use matching software to search for new facilities when they become subject to regulation and to determine whether they already have records somewhere in media programs' databases.

Most of these states have specific data-entry protocols with varying degrees of stringency and enforcement mechanisms so they can control how data for new sites are added to their systems and how core facility data are changed or updated. There are a number of differences, however, in how the states have designed their facility ID systems and how they use their systems to manage and analyze their core records of facility ID information.

Shared Core Facility Data

Six states maintain some type of Master File[42] in their facility ID systems. Their Master Files store a single, legally authorized, and required record of core facility ID information that is used

and shared throughout the agency. These core facility ID elements generally include site name, location, address, owners, and contacts. State agencies that have adopted Master File systems report that the most labor-intensive part of this effort was initially "populating" each Master File. Some states manually reviewed facility files from all their media programs. They then reconciled differences in the media programs' records and entered that information on data input forms to create their Master Files.

In states with Master Files, all the media programs that have a regulatory or other interest in a particular facility or site use the same Master File record for the basic facility ID data. However, each media program still maintains its own media-specific data and any more specialized ID numbers. These numbers are then linked to the Master File, which holds a unique agency-wide ID number for each facility.

State Master File systems ensure that any changes to basic facility ID data will be shared among all state media programs. To keep the Master File accurate and up-to-date, the states have adopted very specific operating procedures that all agency staff must follow when making any changes in shared facility records. Some states place strict limits on who can change facilities' core ID data if they are regulated by more than one media program. In other states, the Master File's software allows various staff to insert, update, or delete facility ID data, but only if they adhere to the agency's specific rules.

Other states allow one media program to change facility ID data on a shared site, but they first require the program staff to contact other bureaus and obtain their comments on proposed changes before they make any edits in the Master File's data. Some states allow trained program office staff to update and change media-specific records for a particular facility, however, as long as those records are not shared by other agency programs and do not require changes in a facility's core ID information.

FRS-type Systems

Six states do not require their media programs to share and use a central facility record. In many of these states, however, the agency has adopted detailed rules that every media program must follow when creating a new record on a facility of interest. These states train their program staffs first to check whether the new facility is already in the agency's database before creating a new facility ID number. When the staffs change existing facility information in one program's database, these states rely on software to reconcile facility ID data with other databases.

Like the FRS at EPA, these applications can automatically flag changes to facility information and pull them into the main agency database, then conduct an after-the-fact reconciliation during a regularly scheduled data clean-up process. During this process, the software automatically updates facility ID records or flags them for a data steward to check them for accuracy. In at least one state, this reconciliation takes place every night. Some states also have rules requiring specific notice to be given to a data steward whenever program staffs make changes in facility ID records.

Benefits of a Central Facility ID System

States report that, by establishing a central facility ID system, they have been able to take a broader approach to permitting, inspections, compliance assistance, and enforcement. Before these states established their central ID systems, their staffs had to conduct lengthy searches of hard-copy files before they launched their inspections just to identify all the permits held by a particular facility. Their facility ID systems enable states to take into consideration companies' entire compliance histories across all media and all facilities with common ownership. This information becomes especially valuable when states are deciding on permit applications or permit renewals because they can identify every regulatory interest at all sites and relate those sites to their respective owners and operators.

States also report that they can respond much more efficiently and accurately to requests for information by the public, the regulated community, and state legislators. They have found substantial savings in staff time and other resources, as well as achieving greater efficiency, more coordinated staff workloads, and more appropriate allocations of other resources. State staff can write their reports more efficiently because data are more accessible and require less time to extract from their computer systems.

States that provide on-line public access to their integrated facility data find that this capability often replaces the need for telephone or written requests for information. Agency staff can then devote their time to high priority work, rather than calls for technical assistance, file searches, and preparing written responses to requesters.

Those states with *shared* Master File systems also have found that, by centralizing their facility records, they now have more accurate and up-to-date facility data. Their media programs are in daily contact with the facilities under their purview, and are thus more likely to have become aware of any recent changes to information. When the media programs discover any changes in basic facility ID data, their agency-wide records are updated in "real-time," eliminating time lags that are more likely to occur when these data are only reconciled or extracted periodically.

Lessons from State Experiences

All states interviewed by Academy staff advised that a successful central facility ID system depends on very strong management, understanding, commitment, and support by an agency's top executives. Early and continued support of agency leaders is essential to create a strong structure for resolving cross-program and cross-organizational issues. In addition, states emphasized that initial commitment of adequate funding to set up and maintain such a system is essential for it to operate successfully.

The states with shared Master File systems also emphasized the need for media programs to cooperate and willingly support the development of a shared central facility ID system. Direct participation and collaboration among media program staffs are necessary for developing a Master File system, for enforcing the rules necessary to operate such a system, and for keeping it constantly updated.

Section Two: EIC's Findings

Finding 1: EPA's media programs do not use the facility records of the FRS in their media-specific databases.

Each EPA media program currently collects and manages the basic identification data for the facilities it regulates. Permit writers, compliance advisors, inspectors, other enforcement staff, and program analysts all conduct their work by relying on the facility information in media programs' separate databases and files. In effect, EPA has multiple homes and "owners" for basic identification information about all of its regulated facilities. Most importantly, the data about a particular facility in one media-specific data system may not match the information in another program's database.

Although some EPA staff and the public use the FRS to facilitate multi-media searches about particular facilities, EPA's media program staffs do not use the FRS's facility records. Instead, FRS creates a separate facility ID that provides merely an interpretation or approximation of the information from the media programs' databases.

If a facility applies for a permit from just one media program, its staff do not have to find out whether that facility already has an FRS record based on data from another regulatory program. Thus, there is no incentive for EPA's program staff to use FRS records, to keep them accurate, or to alert the FRS if they discover that certain facility information has changed.

In any institution, the attention paid to data quality and accuracy depends on whether those data are critical to on-going operations. Unlike the data systems of some state agencies, which specify certain procedures for program staff to follow when using or changing facility ID data, the FRS performs only in the background at EPA, using back-end or after-the-fact reconciliation methods that do not require program staff to maintain its currency or accuracy.

A recent OMB report focused on the confusion created by EPA using a variety of separate identification numbers for each media program and said this problem is a priority for regulatory reforms that will assist the manufacturing sector. However, EPA's response indicated that it would continue to identify separately in each program database the facilities regulated under that program's federal environmental laws, in addition to assigning a facility ID number through the FRS. Although EPA told OMB that it plans to share its FRS numbers with state environmental agencies, EPA apparently does not plan to require that states or EPA's media programs use FRS numbers in existing databases. New facility-level databases created for EPA programs, however, would be required to use the FRS identification number.[43]

Finding 2: EPA's media-specific databases often contain inconsistent identifying data about the facilities regulated by more than one media program; and such basic facility data is frequently inaccurate, out of date, or incomplete.

EPA and others have long expressed concern over the problem of data inaccuracies and inconsistencies in the media programs' data systems.[44] Inconsistent data among the programs'

databases can occur as a result of differences in data formats or in the timing of updates for facility information. As EPA explained in its response to comments on ECHO, data discrepancies among its data systems are often the result of "time differences in reporting, and the fact that different facility staff may report the data in an inconsistent way."[45]

For example, a facility may have reported directly to EPA for the TRI in June 2002, to its state's delegated RCRA program in July 2003, and to a local air quality program when applying for a permit in 2000. The facility's basic identification data may have changed over this three-year period or the facility may by required to report its information in different formats, thus creating inconsistencies among the separate media programs and data systems that all are used for regulating a single facility.

Moreover, some of the facility ID data in the media programs' data systems are either wrong or incomplete because they lack critical information. Environmental researchers who assemble sets of facility-specific data have often found significant gaps in data they have obtained from EPA's program databases. One researcher who surveyed every chemical manufacturing facility in the US using the contact information provided by the program databases found that this information was completely wrong for 35% of the facilities in his survey.[46] Another researcher attempting to link facilities listed in the Permit Compliance System (PCS) to census data found that zip code information was missing from 25% of the facilities with records in PCS.[47]

Finding 3. Inconsistent, erroneous, missing, or outdated facility ID information contained in the media programs' data systems can create misleading linkages or fail altogether to create a link when the FRS is used to aggregate facility data.

While inconsistent or inaccurate facility data in EPA's media-specific databases may not limit the individual program offices when they review permits and conduct inspections, it definitely undermines the effectiveness of FRS in assembling multi-media data about specific facilities from EPA's various program databases. Different or missing facility information can cause FRS's matching software to miss permits or program interests that should all be linked to the same facility. Moreover, it interferes with analyses that program offices might want to conduct, such as comparisons to commercial databases to identify "non-filers" (facilities that should hold permits but do not) and determining the effects of nearby permitted sources on a specific water body.

Changes in facility ID information in any of the programs' databases – such as a change in ownership, splitting up of the property, or a change in the program identifier number – can affect FRS linkages when EPA staff or the public conduct a search to find regulated facilities across EPA programs. When FRS integrates data from EPA's media-specific databases, it magnifies the errors in the underlying program databases. Those errors can have serious consequences when incorrect information about a facility is available to the public for multi-media searches through EPA's on-line databases such as ECHO and Envirofacts.

Comments sent to EPA during the pilot phase of ECHO – which uses FRS records to compile facility data – provide an indication of the magnitude of this problem. Nearly half of the errors (48% of the 6,915 error notifications sent to EPA) "related to facility identification information

(e.g., incorrect facility name or address), permit linkages, duplicates, industrial codes, and other data not associated with compliance or enforcement."[48]   In response, EPA staff explained that these problems are often due to multiple program databases collecting similar but slightly different facility information.[49]

An evaluation of early FRS data by OECA, EPA's primary user of integrated data, also illustrates how these errors can interfere with regulatory efforts. To help identify and target suspected Clean Air Act violators, OECA tried using FRS to compare emission amounts listed by facilities in their annual TRI reports with allowable amounts from the same facilities' permits as listed in the Air Facility System (AFS) database. OECA's first analysis found that 55% (274) of all TRI reporters with over 50,000 pounds of hazardous air releases did not have FRS records that linked to their records in AFS.[50] OECA sent these errors to FRS data stewards, who quickly corrected 143 erroneous linkages.

OECA's second analysis a few months later uncovered another 160 errors in FRS linkages. Facilities that had been linked in OECA's prior analysis were no longer linked in the second analysis. In most of these new cases, FRS had created duplicate ID numbers, which then separated the TRI records from the AFS records.   OECA's third analysis in June 2003 showed 129 facilities reporting toxic air releases to TRI that were not linked to any air permit. While some of the facilities may have gone out of business, OECA found that most showed up as facilities with separate or duplicate FRS records that were incorrectly linked to their TRI reports.[51]

More recently, businesses with multiple installations have found that, when they search their corporate name in ECHO, their name is associated with facilities they have never owned or have not owned for many years.   Companies often find that FRS has assigned more than one number to the same facility. In June 2004, one company performed an ECHO search on several facilities it owns in Alaska and found multiple examples where several FRS numbers had been assigned to the same facility and where the company was identified with a site that it no longer owned or leased.[52]   The recent OMB report mentioned above also identified errors in ECHO searches as the reason that EPA should adopt regulatory reforms to assist the manufacturing sector.[53]

Such errors and inaccuracies in FRS create confusion for the public and costly difficulties for regulated facilities as well as agency regulators. Moreover, community groups and scientific researchers have little confidence in the data they compile from FRS records.

Finding 4:   FRS's back-end reconciliation of facility records is resource intensive, inefficient, and more likely to produce delays and errors.

Currently, EPA corrects or updates its program databases only if facilities notify program offices about the need for changes. To assure consistency among the programs' databases, regulated entities must notify each program office. This process is duplicative and resource intensive, both for facilities and agency staffs.

It is often challenging for companies to file notifications with multiple government agencies when a facility change occurs.[54]   Given the sheer numbers of changes that the private sector

makes every day, discrepancies in program data can quickly accumulate. Businesses shut down facilities or change their facility ID information at a surprisingly high rate. A Dun & Bradstreet study of companies in its global business database found that every 30 seconds 30 new businesses will open, 10 businesses will shut down, 15 companies will change their names, 120 business addresses will change, and 75 business telephone numbers will change or be disconnected.[55]

FRS now extracts new or revised data from EPA's major program databases on a monthly basis. EPA's public databases that rely on the FRS – such as Envirofacts, ECHO, and Window to My Environment – also refresh their FRS data on a monthly basis. Thus, potential errors are likely to arise not only due to inconsistent data in the programs' data systems, but also due to the two-month delay between the time when new information comes into EPA and when it is updated in FRS and made available to the public.[56]

If a regulated entity happens to discover a data discrepancy or error in its FRS records when it performs an on-line data search, it can notify FRS through the automated error notification process. These reports are made, however, only after the information has been released to the public; and the corrections do not necessarily make their way back to the media-specific data systems where they originated.[57]

As a result, data errors identified by companies through the on-line notification process are not automatically corrected in the programs' databases. Even if an FRS data steward notifies a program manager about an error in a facility's ID information in that program's database, some EPA staff have expressed reluctance to change any data unless and until the facility provides formal notice and submits any formal documentation required by EPA's reporting regulations.[58] Thus, facilities must still issue multiple reports or provide documentation for corrections of data errors if they want to ensure that their changes reach all the appropriate EPA programs.

Moreover, the current FRS process unnecessarily consumes the time of EPA's media program staffs and FRS's data stewards, as well as employees of regulated facilities and states. The extra work for FRS's back-end reconciliation approach includes duplicative efforts to update each database whenever facility operations change or new data are collected.

Finding 5: EPA's operational and managerial challenges, rather than lack of technical feasibility, have produced continuing problems with inaccurate and inconsistent facility ID data in the agency's separate media-specific databases.

The EIC's research for this report has revealed that, to provide accurate and timely facility ID data, EPA leaders must overcome the agency's strong cultural bias that favors the media programs' separate databases and opposes efforts to integrate those systems. During more than a decade, there have been many calls for EPA to eliminate these operational and management problems and establish a more integrated approach to its information management. But EPA continues to maintain its separate media-specific data systems and has not been able to overcome the agency's bias against data integration.

As several states have recently demonstrated, technology is no longer an obstacle to integrating

media-specific data systems. State agencies that have adopted centrally managed facility ID repositories (or Master File systems), which they share agency-wide, readily attest that one of the most challenging issues is not technical. Instead, the main obstacle is obtaining acceptance by program managers and staff to share a single Master File system of facility ID data so they will commit to the necessary processes for entering data into that system.[59] Many state agencies have tested and proven the technological feasibility of a single, shared facility ID system and have demonstrated its value in terms of efficiency, effectiveness, and reduced costs.[60]

In short, the EIC finds that technical feasibility is not an obstacle to EPA establishing a single, shared Master File system of facility ID data. EPA's much larger challenge is overcoming its operational and management difficulties, especially the stovepiped single-medium culture that pervades the agency.

Section Three: EIC's Recommendations

1. EPA should develop, implement, maintain, and oversee a single Master File system for all core facility identification (ID) data and should begin this process as soon as possible.

EIC members believe that, as soon as feasibly possible, EPA must establish a single shared national database of core facility ID information in order to integrate and manage effectively all facility-specific data from throughout the US, so EPA and the states can meet the needs of the many parties that collect, report, and use environmental information. This Master File system should contain the officially authorized ID records for each regulated facility that all of EPA's media programs must use.

For EPA to establish and maintain a single shared Master File system, it must adopt a different approach than it has used when developing its own and state databases during the past 30 years. This new approach will require that EPA shift to managing basic facility ID data on an agency-wide basis, rather than collecting and maintaining multiple sets of such information in the media programs' separate databases. This new approach will also require collaboration among state agencies, EPA's media programs, and its regional offices on developing, managing, and updating the shared Master File system now and for the foreseeable future.

Commitment to Support

EPA's success in accomplishing the fundamental changes necessary to operate this Master File system will depend on strong commitments and support from the agency's top leaders, the OMB, and Congress.

The EIC has found that EPA's greatest challenge to establishing a single shared Master File system is operational and managerial, not technical. To overcome this hurdle, EPA's Administrator must take several actions:

- Exercise clear leadership and strong commitment to integrating the agency's data systems

- Maintain on-going and sustained attention to integrating those systems and overcoming the agency's stovepiped culture

- Require that the separate media programs cooperate on establishing and sharing a single Master File system for basic facility ID information

- Designate a key agency official and give that person clear authority to implement the shared Master File system

- Provide adequate financial resources and staff with appropriate technical expertise to implement and maintain the Master File system

The Administrator will find strong support for these actions throughout the regulated community, the states, environmental groups, Congress, and the general public.[61] The enormous value of these changes will be clear as EPA's data systems become more efficient, effective, and equitable. They will save time and money for all parties; they will increase data accuracy and make it easier to identify cross-media environmental or public health problems; they will be more fair in reducing the reporting burden of the regulated community, especially small businesses; and they will help EPA and state agencies to address cumulative risks where the cross-media environmental and public health threats have not previously been identified, such as in low-income or people-of-color communities.

Commitment to Fund

This new comprehensive and integrated approach for collecting and managing facility ID data will also require a commitment of sufficient and continuing funds to assemble the Master File system and to update it continuously. Resources to design, implement, and maintain an integrated Master File system will be needed for:

- Application development, including eventual upgrades and maintenance;

- System planning, such as documented business rules, common data definitions and standards, user operating procedures, data dispute resolution procedures;

- Human resources, including staff time, management and staff training, data entry, data reconciliation from legacy systems, quality assurance, correcting data errors by agency staff and regulated entities, information management, and technical support;

- Identification and development of potential rule changes.

Major information technology investments are currently being made at both state and federal levels. EPA's FY 2005 budget includes $25.0 million for state grants to develop nodes or web servers and upgraded systems that will facilitate the ability to interface between their databases and EPA through the Exchange Network.[62] During these developmental stages, it will be easier and less expensive to institute the changes necessary to coordinate with a Master File system, along with its rules and standards for implementation. EPA should take the lead now in developing its Master File system for integrating facility ID data, and should mandate the use of uniform data standards and definitions, while states are still in the process of developing their own upgraded systems.

As the EIC has learned from state agencies that have implemented a Master File system, opportunities also exist for substantial cost savings. For example, a shared Master File can reduce staff workload by reducing the time needed for quality assurance. Even more important, where states have created their Master Files as an integral part of modernizing their data systems, they can support inspectors in the field, write permits and enforcement actions more efficiently and quickly, simplify and reduce reporting costs for regulated entities, analyze reported data more fully and fairly, notify the public more quickly and accurately, reduce the costs of

responding to public inquiries and FOIA requests, see geographic relationships between stressors and vulnerabilities, and see patterns across and within programs and industrial sectors.[63]

A single Master File system shared by all of EPA's media programs and the states will also afford benefits and cost savings to businesses and other regulated entities. They can reduce their employees' time for filing reports, make fewer data corrections, and provide more frequent updates, as well as gain increased accuracy and timeliness of facility information available to regulators and the public.

As many states with such shared data systems have learned, high-profile agency leadership and advance commitments of adequate funding to build and maintain their systems are essential ingredients. These two factors will be key to EPA's success in integrating facility records, saving agency costs, providing accurate and timely information to the public, and reducing the reporting burdens of regulated businesses.

2. The media program offices at EPA and the state environmental agencies should collaborate in planning, designing, and coordinating how the single Master File system will operate, including (a) development of uniform procedures and definitions, (b) how the system will track facilities' identities through changes in name, operations or ownership, and (c) how data elements that identify and define each type of facility can support the needs of the various media programs and the state agencies.

Establishing a collaborative effort to decide on details about the system's design and operation, uniform procedures and definitions, and the agencies' respective roles is one of the most important threshold matters that EPA must first address to ensure the success of an integrated national facility ID system. One of the initial decisions will be to determine how to administer the Master File records.

The Master File system could be designed to operate in one of three ways:

(1) EPA could administer the records in the Master File system through a central EPA office and require that all media programs use that system as the authoritative source for ID data on facilities that are subject to federal regulations;

(2) Individual state environmental agencies could administer the records in the Master File system that relate to facilities within their jurisdictions, and their facility ID data could then be submitted to EPA or pulled into EPA through the Exchange Network to assemble the national Master File system; or

(3) Similar to the usual process for EPA approval of states that qualify to assume responsibility for administering federal environmental requirements, EPA could authorize states to administer the Master File records for facilities within their jurisdictions as long as they agree to follow established uniform procedures and data standards.

For any of these approaches, EPA alone would manage and update Master File records on facilities that are regulated only at the federal level.

While all three designs are technically feasible, the EIC members emphasize that, to assure accuracy and consistency of ID data in the Master File facility records, whoever administers the system and maintains those records must adhere to a uniform set of procedures and definitions of data elements.

For example, uniform procedures or business rules must determine:

- Who maintains the core facility ID data
- Who can change those data
- How the information will be changed
- What procedures to follow when creating a record for a newly regulated facility

As described above in Section One, EIC members have found that states vary in the stringency of their rules for managing additions and changes to their Master File systems. EPA should build on these state rules by working closely with state agencies and its own media programs to develop procedures to ensure that the Master File collects and maintains correct and current data, while avoiding interference with the media programs' routine work.[64] EPA can take advantage of lessons learned from the states when developing and implementing its own business rules for sharing the Master File system.

Additionally, agreement on standard definitions of data elements for facility records in the Master File is also critical. The definition of "facility" is a particularly challenging issue. For example, in establishing a facility record, it must be determined whether to define a large complex that contains a variety of operations or different equipment, such as a refinery or military base, as one or several facilities. A confined animal feeding operation (CAFO) may be defined as a 10-animal operation in one state, while others may only consider 1000-animal operations as CAFOs. If different sizes of facilities are labeled as CAFOs in the Master File system, it will confound users' ability to understand the magnitude of the environmental impacts or to evaluate the effective management of those impacts.

While it might seem easy to determine the address or data element for the location of a facility, identifying the location of regulated entities with unique profiles often becomes problematic. For example, a mobile drilling rig may change its location; or two dentists' practices on the same floor of the same building each may require RCRA permits, while their office building may be required to have an air permit. Although their physical location may be the same, they may be distinct entities; or the regulated facility may stay the same, but its location may change.

In addition, state agencies or EPA's program offices may record "ownership" of facilities differently. Some may record both current and historic owners, while others record only the current owner. For the Master File system to find all facilities owned by specific companies, the data must consistently determine and record their ownership.

EPA and the states should collaborate to address these differences in definitions as they develop the Master File system. During this process, they can also identify potential areas where existing

statutory definitions, either state or federal, may need to be reconciled so that shared facility ID data can accommodate all the regulatory requirements associated with regulated facilities.

3. The EPA Administrator should direct all EPA media programs and regional offices (a) to use and share the single Master File system as EPA's <u>only</u> source of core facility ID data and (b) to assure that future updates or other changes to facility ID data can be directly submitted by facility managers to the Master File system in accordance with procedures jointly designed by EPA and the states.

Once the single, shared Master File system is developed and established, it is imperative for its continued accuracy and viability that EPA's media programs and regional offices use it. As one state representative explained, core facility ID data only remain accurate and current if program staffs "touch" the facilities' shared Master File records almost every day.[65] In addition, it is imperative that facility managers submit changes to their own facilities' Master File data directly to the Master File itself, within the constraints of regulatory requirements, rather than indirectly to other multiple and decentralized systems.

Specific, disciplined, and uniform procedures for using the shared Master File system will also help to keep program staff from trying to make "end-runs" by not using the central ID data. For example, in one state system, the permitting division chief refuses to approve any documents that permit writers have not prepared using the Master File system. This point of control keeps program employees from by-passing facility records in the Master File, and they cannot generate new facility ID data without first using any information that is already in the shared system.[66] Another state is using a carrot-and-stick approach to reward with special recognition and promotions for its staff who demonstrate early leadership by using or updating the Master File's facility ID data.

One way to assure continued use of the system is to make it useful for media programs' routine work and to offer links to media-specific data about facilities. While certain core facility ID information can be shared, media program offices also need various more specialized data about facilities. For example, a single facility may have more than one data element for the outlets of its discharges or emissions. EPA's and states' water programs have more interest in the locations of outfalls or discharge points than in the facility's street address, and other media programs may need different mailing addresses to contact a facility about particular regulatory issues.

Data in the PCS may relate to laboratories used by each facility to document water sampling data, while annual TRI reports might be submitted by the company's law firm, and RCRA data might be reported by the company's headquarters. Some media programs, like hazardous waste, need facilities' historical information and need to store that information as separate records in the data system even after a facility closes and is no longer regulated by other programs. Researchers may also need sets of historical data so they can compare data from the same facilities over various time periods.

EIC members heard concerns from the staff of some EPA and state media programs that a shared Master File would threaten their ability to collect and directly control facility-specific data. In

the course of this study, however, EIC members have also heard from other EPA and state program managers who support a shared Master File system as long as it also can link to or incorporate the media-specific information they need to manage routine tasks. They also want their media programs to continue "owning" the data uniquely related to their regulatory requirements.

As demonstrated by several states currently operating shared Master File systems, technology now can support systems that incorporate multiple sets of facility ID data for each media program. States with newer data systems also can provide different sets of contact information as needed by their media programs, while still maintaining a common, shared set of core facility ID information that allows the Master File system to integrate all the environmental data about a facility from the various media programs.[67] These states thus demonstrate how unique facility ID numbers contained in their Master File systems can also provide staff with access to underlying information about a particular facility that may be of interest to each media program.

4. In collaboration with state environmental agencies, EPA should review and modify as needed all the relevant reporting rules or other requirements for its media programs, regions, and state program delegations so the single Master File system is recognized as the official legal record of core ID data for federally regulated facilities; EPA will thereby legitimize use of facility ID data in the Master File system to satisfy all federal data collection and reporting obligations.

Changes to EPA rules or other requirements that will allow federally regulated facilities to report their ID data to the shared Master File as their official legal record would accomplish two important goals of data integration: greater data accuracy and reduced reporting burdens for businesses.[68] Current environmental regulations often require facilities to submit their ID data separately to EPA or state agencies for each media program.[69] When facilities need to change their core ID data, they must file reports with multiple agencies and multiple media programs.

Once it is established, the shared national Master File should be designated as EPA's official legal record of each federally regulated facility's core ID data, so that all EPA and state media programs and enforcement offices can rely on the information in the Master File's records. By legally recognizing this integrated data system, EPA would motivate regulated facilities to keep their ID records updated and accurate and would ensure that any changes are uniformly disseminated to all regulatory staffs and other interested parties.

In return, this change would lessen the reporting burdens on regulated businesses and would tend to encourage further electronic reporting. Facilities would no longer need to submit multiple forms notifying each EPA media programs about changes in their ID data.

Rule changes establishing that ID data in the Master File system are EPA's legal records for each federally regulated facility will require a significant, but certainly manageable, commitment of time and resources. States would also need to examine any of their reporting requirements based on EPA's rules to see if they need to make changes too. It is beyond the scope of the EIC's work to determine exactly how EPA and the states should adopt these changes. However, the EIC

recommends that EPA immediately initiate a joint effort with the states to ascertain the specific policy, regulatory, and legal changes that may be needed.

The ways that EPA and the states could make these changes may include:

(A) A single rule for all media programs that would provide the legal framework by allowing facilities to report the ID data to the Master File system, for the purposes of all EPA and state environmental programs as a substitute for media-specific reporting requirements; or

(B) Separate rule changes for the reporting requirements of each of the EPA and state media programs.

Regardless of how EPA and the states choose to revise their rules, changes in these requirements are fundamental to establishing a truly integrated system of multi-media environmental information that is linked through a single, shared Master File containing the official ID data for regulated facilities.

5. EPA should require all its media programs and regional offices, as well as state environmental agencies, to adopt and use uniform data elements and standards that identify all regulated facilities, based on the common facility ID elements, definitions, and standards that have been jointly adopted – and will continue to be updated – by the State-EPA Environmental Data Standards Council.

The requirements for adhering to common data standards and agreeing to the format and definitions of these data elements are critically important for establishing one record of core facility data that all users can collect, share, exchange, and update. The core data elements in the single, shared Master File system should consist of the same basic facility ID information that is common among the various states and EPA's media programs. They will likely include name, address, location, ownership, and central contact information. [70]

Much of the work to develop facility ID data standards has already been done, but it is an iterative process, and EPA can build on the efforts of the EDSC and State-EPA IMWG. These groups should be expanded to include representatives from local and tribal agencies in the development of common data standards. Given the importance of this information to the public, the regulated community, regulatory agencies, and others, EIC members further urge this collaborative group to disseminate draft data standards broadly so that all affected parties can submit comments before the group adopts final standards.

The most difficult step, however, is not developing data standards but ensuring that the media programs and the states use them. Through the Office of Environmental Information, EPA's Administrator should continue to exert control over the agency's funding for information technology to influence and require that all media programs and states adhere to uniform data standards.

6. EPA should provide assistance to state agencies in adopting and implementing those facility ID standards and in developing the states' capacity to maintain and update their own facility ID records so they can operate effectively in conjunction with the single Master File system.

Concurrent with a requirement that state agencies follow data standards, EPA should provide technical and financial assistance for the states to make these changes. Just as the development of EPA's shared Master File can benefit from the lessons learned from the states' newer facility ID systems, state agencies that do not have such systems can learn from the Master File as they develop their own integrated systems to operate in conjunction with the national one. State systems could include records for facilities that are only regulated by the states as well as those regulated by an EPA program, and the states' capacity to integrate their data with EPA's data will make it possible for all agency managers to conduct broader and more comprehensive analyses of their environmental information.

7. Every year, EPA should evaluate the accuracy of facility ID data in the single Master File system and then report its findings annually to the states, the public, regulated facilities, and Congress so the system can be continuously improved.

Ensuring that the shared Master File system contains high quality data is critical for building trust among all the system's potential users, especially EPA's media program staff. This trust is essential for a single, shared Master File system to work effectively at EPA. Use and acceptance of the shared system throughout EPA and all state agencies, as well as by regulated businesses, the public, and researchers, will only happen if all users adhere to uniform data standards and procedures, as well as strict requirements for data quality.

On an annual basis, EPA should evaluate and report publicly how well its own media programs and the state agencies have followed the procedures and data standards. This report should include statistical data on the specific rates of data accuracy and the rates of responsiveness to correction requests, broken down by each media program, state, and industry sector or type of facilities.

8. EPA should provide public access to its single Master File system through an on-line, web-based interface that is easy for the public to understand; and that interface should provide for public access to agencies' information about the environmental obligations and performance of all federally regulated facilities.

EPA's on-line search applications currently rely on the FRS to enable public searches of the agency's major databases. Under the current system, the public sometimes receives inaccurate information about specific facilities. EIC members believe that, once EPA fully implements the single Master File system recommended in this report, it will allow the public to conduct much more accurate, productive, and up-to-date searches through EPA's web site.

Because the Master File system will promote adoption of uniform data elements and standards, it will enable EPA and the state agencies to develop and offer more sophisticated, value-added search capacities. For example, Master File data could be organized and analyzed to determine

which facilities in a certain watershed hold permits allowing them to discharge a particular pollutant. Searches for this type of information are currently difficult to perform and, when attempted, the accuracy of their results is quite uncertain.

EPA is probably best positioned to manage a web site that offers this multi-media search capacity to the public because it would have access to information about facilities across the country. Thus, EPA could provide the browser and other analytical software to combine the requested data in easily usable formats and provide links to individual states' web sites for even more detailed or comprehensive information. EPA's role in collecting the data from all facilities, organizing it, and disseminating it is essential to make this information usable by both the public and government agencies.

Moreover, state agencies and EPA's media programs or regional offices can derive added value from the Master File's information through better analysis and comparisons once the data are available in a single location. As a result, their staffs will be more likely to modify their procedures and operations to work in conjunction with the data in the Master File records.[71] But to offer this benefit and provide this function, EPA does not necessarily have to hold or control all the state agencies' or programs' data on regulated facilities. Using its new AIRNow web site as a model, EPA could instead use the shared Master File system to integrate, aggregate, and organize particular sets of data from all EPA and state data systems so they are always current and are more readily accessible, usable, comprehensible, and comparable.[72]

Section Four: Conclusion

The fragmented nature of environmental statutes, rules, and programs has made it difficult, if not impossible, for EPA to integrate environmental data across media programs.

Exemplary efforts by EPA and the state environmental agencies to adopt data standards and begin sharing their information with help of newer technologies – particularly over the last three years – have begun to overcome these hurdles and have demonstrated the potential for more effective data integration. But these efforts are still only nibbling at the edges and do not reach the heart of the problem – the fragmented interactions among media programs and regulated facilities that, in turn, deprive the public, state and EPA regulators, and businesses from access to multi-media information about environmental impacts.

EIC members all believe strongly that EPA and states lacking upgraded data systems need to revamp the way they manage environmental information. What the US needs is a single, shared Master File system for facility ID data. That system can then be shared nationwide and can provide all parties with the official legal records for every federally regulated facility, regardless of the media program or programs that regulate it.

Drastic actions are not necessary to adopt and manage a shared Master File system. EPA and the states are already making major investments in new or upgraded systems. Some states have already developed their own integrated Master File systems. Other states are eager to adopt such systems soon, but they want guidance so they can ensure their plans are compatible with EPA's data systems. Now is the time for EPA to move quickly on reforming its technology and rules for facility ID data so EPA and the states can proceed with true integration of the nation's environmental data.

Although EIC members represent the interests of very different parties, this report reflects their strong agreement about the urgent need for EPA to adopt a consolidated system for collecting, sharing, and updating facility ID information. EIC members hope that their findings and recommendations in this report will help to push Congress and EPA toward the structural changes necessary to provide a truly effective system for managing our nation's wealth of environmental information.

ENDNOTES

[1] *Using Information Strategically to Protect Human Health and the Environment: Recommendations for Comprehensive Information Resources Management: Report of the IRM Strategic Planning Task Force* (Washington, D.C.: August 1994), final recommendations reprinted in *National Advisory Council for Environmental Policy and Technology (NACEPT): Past and Future, A Decade of Stakeholder Advice* (Washington, D.C.: July 1999) Appendix I-7. To implement this recommendation, the Task Force urged EPA to "develop, immediately implement, and enforce data standards; develop data integration policies and tools; define data requirements and identify gaps in the data inventory; and reduce the burden on providers of information." The Task Force members consisted of representatives from state, local and federal agencies, public and environmental interest groups and an academic institution.

[2] EPA reiterates the importance of environmental data in its 2003-2008 Strategic Plan: "Accurate, timely, and usable information is the foundation of decisions and actions taken by EPA, states, and others responsible for protecting human health and the environment." Environmental Protection Agency, *2003 – 2008 EPA Strategic Plan—Direction for the Future* (Washington, D.C.: September 30, 2003),134.   See, e.g., National Academy of Public Administration, *Environment.gov: Transforming Environmental Protection for the 21st Century* (Washington, D.C.: June 2001), 170.

[3] See e.g., National Academy of Public Administration (NAPA), *Evaluating Environmental Progress: How EPA and the States can Improve the Quality of Enforcement and Compliance Information* (Washington, D.C.: June 2001), 22-26; NAPA, *Environment.gov*, 171; Environmental Council of States (ECOS), *State Environment Agency Contributions to Enforcement and Compliance* (Washington, D.C.: April 2001), 14; U.S. General Accounting Office, *Major Management Challenges and Program Risks: EPA*," Report #GAO-01-257 (Washington, D.C.: January 2001), 17-18; NACEPT, *Final Report and Recommendations of the Information Impacts Committee: Managing Information as a Strategic Resource (Washington D.C.: January 1998); Facility Identification Initiative; Notice,* 61 Federal Register 52587-52600 (October 7, 1996); 1994 NACEPT Task Force, *Using Information Strategically;* U.S. Environmental Protection Agency, *Strengthening Environmental Management in the United States: Report of the Task Force to Enhance State Capacity* (Washington, D.C.: July 1993), 18.

[4] Environmental Council of States (ECOS) and U.S. EPA, *Environmental Pollutant Reporting Data in EPA's National Systems: Data Collection by State Agencies* (Washington, D.C.: September 1999), 1.

[5] NAPA, *Environment.gov,* 171, NAPA, *Evaluating Environmental Progress,* 22. U.S. General Accounting Office, *Major Management Challenges and Program Risks: EPA*, January 2001.

[6] Most of these changes in environmental protection strategies have been encouraged or mandated by Congress, such as the Pollution Prevention Act and the Government Performance and Results Act of 1993. These new approaches require a change in environmental data systems "capable of monitoring, collecting, integrating and analyzing the various kinds of data across all environmental media in order to ensure that performance-based management can operate successfully." NAPA, *Evaluating Environmental Progress*, 7.

[7] See, e.g., "Summary of State Responses about State Facility Identification Systems," Appendix B.

[8] See, e.g., Coalition for Effective Environmental Information, *Comments on the Enforcement and Compliance History Online Project*, submitted to EPA on March 31, 2003; NAPA, *Evaluating Environmental Progress*, 23.

[9] Environmental Protection Agency, Information Utilization and Targeting Branch, Office of Compliance, *Enforcement and Compliance History Online: Public Comment Period Summary and Response to Comments*, Aug. 19, 2003, 12-13. In its response to comments on the ECHO pilot, EPA acknowledged that "[t]he pilot ECHO use indicates the multiple collection of similar facility data is cause for concern" and that this inconsistency was the major factor in the facility identification errors found during the pilot.

[10] See, e.g., Conoco-Phillips Alaska Facilities in ECHO (from 6/30/04 Download), as discussed in Finding 3 and endnote 52 of this report.

[11] Coalition for Effective Environmental Information, *Comments on the Enforcement and Compliance History Online Project*, submitted to EPA on March 31, 2003, 4.

[12] National Federation of Independent Businesses, *National Small Business Poll: Paperwork and Record-keeping*, Vol. 3, Issue 5 (2003), 8.

[13] "Testimony of Carol Andress, Environmental Defense Fund, Before the Committee on Small Business on Reducing Environmental Regulatory Burdens on Small Business," (March 7, 1996), 2.

[14] Id. For the project, the Environmental Defense Fund (EDF) sponsored the development of software that would integrate the various reporting requirements, provide information on the requirements in a more understandable way, provide links to pollution prevention alternatives, and help educate the industry on environmental and human health impacts on its operations.

[15] For a discussion on other ways multi-media analysis of facility information can be used, see U.S.EPA, Office of Environmental Information, *Toxics Release Inventory Division, How are the Toxics Release Inventory Data Used—government, business, academic, and citizen uses* (May 2003).

[16] See e.g., Information Management Workgroup and Environmental Health Data Action Team, *Primer on Environmental Public Health Information Activities*, July 2004. As an example of one of the many public health information activities now being pursued, see the Florida Data Linkage Demonstration Project, which is attempting to link EPA's TRI data with statewide surveillance systems for asthma, cancers, and other illnesses. http://www.cdc.gov/nceh/tracking/EPHTracking/contacts/fl.htm

[17] "Corporate Environmental Behavior and the Effectiveness of Government Interventions, Proceedings of Lunch Panel Discussion: Progression Towards an Environmental Facility Research Database," April 26-27, 2004, sponsored by the U.S. EPA National Center for Environmental Economics and the National Center for Environmental Research, 1.

[18] "Corporate Environmental Behavior and the Effectiveness of Government Interventions, Proceedings of Lunch Panel Discussion: Progression Towards an Environmental Facility Research Database," (April 26-27, 2004), sponsored by EPA's National Center for Environmental Economics and the National Center for Environmental Research, 10.

[19] The first facility identification system used by EPA, known as the Facility Index Numbering Data System (FINDS), began with an initiative developed in Region 2 in the early 1980s. The Facility Linking Application used the FINDS numbers to produce the linkages.

[20] IDEA provides an historical profile of inspections, enforcement actions, penalties assessed, toxic chemicals released, and emergency hazardous spills for EPA regulated facilities. www.epa.gov/Compliance/planning/data/multimedia/idea/index.html

[21] U.S. Environmental Protection Agency, Office of Information Resources Management (OIRM) for the Facility Identification Initiative, *Analysis of Data Integration Costs and Policy Options for Improving the Quality of EPA's Integrated Facility Data* (May 9, 2000), 2.

[22] Ibid., 2.

[23] In one example, approximately 35% of the facilities listed in the heavily reviewed Sector Facility Indexing Project (SFIP) were found to have errors in the core facility identification data using the matching process in the Facility Linking Application. Ibid., 2.

[24] "Back-end integration creates 'virtual links' between existing, separately-managed facility files." Ibid.

[25] Ibid., 40.

[26] Ibid.

[27] The initial method for matching may vary depending upon the data system being scanned. For example, the Risk Management Program database compares the latitude and longitude lines of a facility to find out whether the data submitted is from the same facility. For the National Emissions Inventory and Toxics Inventory submissions, the FRS uses the "contact person" to compare facility information and to reconcile which submissions come from the same facility. Interview with OEI staff (January 2003).

[28] These elements are set forth in conformance with the data standards established by the Environmental Data Standards Council for these elements and compliant with the guidelines in the Federal Identification Template for States.

[29] If the facility does not report to the TRI, the FRS will then go to the RCRA program database, and so on, in a specified descending order. Interview with OEI staff (November 17, 2004).

[30] An update frequency table is listed at http://oaspub.epa.gov/enviro/data_update?

[31] U.S. Environmental Protection Agency, *US Environmental Protection Agency Into Action: Enterprise Architecture Status Report 2004 (*September 13, 2004), p.1. EPA received a "Green" rating from the Office of Management and Budget (OMB) in both status and progress for the Agency's implementation of the E-Government initiatives on the President's Management Agenda (PMA) Scorecard, reflecting success in meeting requirements for its enterprise architecture, among other activities. See, Office of Management and Budget, *The Federal Government is Results-Oriented: A Report to Federal Employees* (August 2004).

[32] The FRS is the underlying search mechanism for EPA-sponsored web services such as: 1) Enviromapper, http://www.epa.gov/enviro/html/em/index.html, which provides interactive maps that help users visualize environmental information and demographic data in local settings, at regulated facilities, Superfund sites, and other areas of interest,; 2) Envirofacts, http://www.epa.gov/enviro/, which provides access to several EPA databases from which a user can obtain information about environmental activities and regulated facilities, such as what facilities have permitted emissions releases or which public water systems have reported violations to EPA, that may affect air, water, and land anywhere in the United States; 3) Window to My Environment, http://www.epa.gov/enviro/wme/, which provides geographic statistics about any particular area of the country, including population, the location of water monitoring sites, watersheds, and provides links to state and local data on air and water quality, as well as other environmental conditions; 4) Enforcement Compliance History Online (ECHO), http://www.epa.gov/echo/, which provides compliance information on regulated facilities; 5) Integrated Data for Enforcement Analysis (IDEA), http://www.epa.gov/echo/, a multi-media search tool for obtaining environmental performance data on regulated facilities, and AIRNow, http://www.epa.gov/airnow/, which provides public access to air quality information all over the country. A recent addition to EPA's System of Registries, the Registry of EPA Applications and Databases (READ), provides a comprehensive list of all applications owned by EPA's offices and regions. www.epa.gov/read

[33] EPA's Integrated Error Correction Process (IECP) can be found throughout the EPA web site, and is available from EPA's Home Page as well as sites such as Envirofacts, ECHO, the Pesticide Data Submitters List, AIRSData, and others. EPA developed its IECP to comply with guidelines set forth by the Office of Management and Budget at 67 Federal Register 8451 (February 22, 2002). EPA's guidelines, "*EPA's Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency*" can be found at: (http://www.epa.gov/quality/informationguidelines/documents/EPA_InfoQualityGuidelines.pdf

[34] The states in the Environmental Council of States (ECOS) and EPA jointly formed the Information Management Working Group in 1998 and committed themselves to a partnership to build locally and nationally accessible, cohesive and coherent environmental information systems.   http://www.epa.gov/oeiinter/imwg/index.htm

[35] The FITS documents prepared by joint EPA/state groups provide guidelines to help states identify common data elements to enable the sharing of facility data and to help states succeed in the integration of facility site information. Facility Identification Template for States, Version 2 (FITSII), sponsored by Knowledge Transfer Action Team of the State EPA Information Management Workgroup, Washington Department of Ecology, Environmental Council of States, and US EPA One Stop Reporting Program.

[36] The Environmental Data Standards Council (EDSC) is made up of 10 members from EPA, states and tribes. http://oaspub.epa.gov/edr/edsc$.startup

[37] The Facility Identification Data Standard consists of core data elements that properly identify the location, the affiliated organizations, individual business activities, and the environmental interest of a facility site. The twelve final standards adopted by the EDSC are: biological taxonomy, chemical identification, contact information, date, enforcement/compliance, facility site identification, federal facility identification, latitude/longitude, permitting information, reporting water quality results for chemical and microbiological analytes, SIC/NAICs, and tribal identifier. http://www.envdatastandards.net/section/standards/approved/

[38] See http://www.exchangenetwork.net/

[39] Five of these exchanges involve data traditionally shared between the states and EPA such as information on air emissions, water permits/compliance, drinking water quality, hazardous waste handlers, and facility identification. The remaining six exchanges are entirely new flows of data among regional and state partners or between the states and EPA, such as beach monitoring, ambient water quality, hazardous waste shipping, and ambient air monitoring. "Network Update," January 7, 2005. Environmental Information Exchange Network, http://www.exchangenetwork.net/progress/progress_chart.pdf.

[40] Interview with OEI staff (November 17, 2004).

[41] Locational Information Improvement Project, http://www.epa.gov/enviro/html/locational/ldip/index.html. EPA states that "[t]he primary objective of this effort is to identify, collect, verify, store, and maintain an accurate, consistently documented set of locational data for entities of environmental concern. A secondary objective is to support the infrastructure needed to manage these data in a manner that yields integration across national, regional, tribal, and state systems. The intent is to support EPA's movement toward data integration based on location, thereby promoting the use of EPA's data resources for a wide array of cross-media analysis, such as community-based ecosystem management and environmental justice."

[42] States identify this record of core facility data shared agency-wide in different ways, such as Site Master File, Facility Master File (FMF), Client and Site information, or Core information. Throughout this report, the EIC will refer to this shared record of core facility data as Master File.

[43] Office of Management and Budget, Office of Information and Regulatory Affairs, *Regulatory Reform of the U.S. Manufacturing Sector: A Summary of Agency Responses to Public Reform Nominations*, 2005, 15. As part of a government-wide effort to reform regulation of the regulated community, OMB solicited nominations for specific reforms that could result in lower costs or greater effectiveness, among other things. In this report, OMB identified what it considered 76 "priority reforms" from the 189 nominations, one of which was a complaint by Deere & Company over multiple identification numbers in the EPA systems.

[44] NAPA, *Environment.gov,* 171; NAPA, *Evaluating Environmental Progress*, 23; NACEPT Task Force, *Using Information Strategically*; *Facility Identification Initiative*, 52589; US EPA, *Analysis of Data Integration Costs*, 8; GAO, *Major Management Challenges*, 17; ECOS, *State Environmental Agency Contributions to Enforcement and Compliance*, 55.

[45] "ECHO Public Comment Period Summary and Response to Comments" (Aug. 19, 2003), prepared by the Information Utilization and Targeting Branch, EPA's Office of Compliance, 37. http://www.epa.gov/echo/info/echo_review_period.pdf

[46] "Corporate Environmental Behavior and the Effectiveness of Government Interventions," Proceedings of Lunch Panel Discussion: Progression Towards an Environmental Facility Research Database," (April 26-27, 2004), 10.

[47] E-mail correspondence with an Associate Professor at the Department of Agricultural and Resource Economics, University of Maryland (July 28, 2004).

[48] "ECHO Public Comment Period Summary and Response to Comments," Aug. 19, 2003, prepared by the Information Utilization and Targeting Branch, Office of Compliance, EPA, 12. During 2004, error notifications submitted to Envirofacts and ECHO averaged 87 and 106 per month, respectively. E-mail correspondence with OEI staff (November 9, 2004).

[49] Ibid., 13 and, 37.

[50] This information is based on an evaluation prepared by EPA's Office of Environmental Enforcement and Compliance Assurance (January 2003).

[51] An evaluation of the FRS linkages between program databases is also possible using data from OECA's Sector Facility Indexing Project (SFIP). The SFIP provides the public with facility-level profiles for five major industry sectors, including compliance and enforcement records and information about the manufacturing processes. Before the information is released to the public, EPA allows the facilities an opportunity to review the data to assure the correct universe of facilities in each sector and the accuracy of the data. http://www.epa.gov/sfipmtn1/index.html. An analysis performed in 2002 revealed that the 890 facility records in the SFIP produced 1119 records in the FRS, indicating that some single facilities were listed more than once in FRS.

[52] Conoco-Phillips Alaska Facilities in ECHO (from 6/30/04 Download). For one North Slope site EPA had assigned four FRS numbers to the same facility. In another example, EPA had assigned one FRS number to what were, in fact, three distinct facilities. In addition, while the company had notified EPA in 1995 about the sale of another facility, the FRS system still showed ConocoPhillips as the facility's owner. Similarly, FRS shows ConocoPhillips as operating a facility at a site that it once leased in Fairbanks. In fact, Alyeska Pipeline now leases that property, and Alyeska has received a different FRS number from EPA for that same site. In August 2004, ConocoPhillips provided EPA's OEI with the information from this download to make the specified corrections in the FRS records, which EPA agreed to do.

[53] OMB, *Regulatory Reform of the US Manufacturing Sector*, 16.

[54] In its reply to comments on ECHO received by OMB, EPA has recognized the "challenges [for regulated entities] of reporting this information at multiple times to multiple offices." EPA Office of Compliance, *ECHO Public Comment Period Summary and Response*, 29 "The pilot ECHO use indicates the multiple collection of similar facility data is cause for concern. Facilities should ensure that facility name changes are reported to the proper regulatory authorities as required under the law. This will ensure that consistent information is shown on the ECHO report. A great number of notifications sent to ECHO were the result of facilities failing to notify regulatory agencies of recent corporate name changes, or because a company preferred an alternate name or address." EPA Office of Compliance, *ECHO: Public Comment Period Summary and Response*, 13.

[55] Tom Marsden, Dun & Bradstreet and Reginald Berry, FairIsaac, "Preventing Improper Payments: Leveraging Data & Risk Modeling Solutions," September 2004, p. 8. While EPA is now working with Dun & Bradstreet to receive its updates to changes in company ownership, these updates do not make their way systematically to the program databases.

[56] Interview with OEI staff, (November 17, 2004).

[57] As the Academy noted in a 2001 report, systems that aggregate data on regulated facilities "should be based on accurate data and should not, as a standard operating procedure, rely on regulated entities to identify mistakes only after the data have been released to the public." NAPA, *Evaluating Environmental Progress*, 23.

[58] Discussions with EPA staff. This reluctance stems from a perceived impact on enforcement actions where intent or negligence is an issue if contact information changes are made without required documentation. The EIC has not performed a legal analysis of this issue to determine what impact using a central source of contact information for a facility's identification record would have on enforcement needs. One state has avoided inconsistency between the central registry data and the program database data by asking the facility to provide formal notification to the program offices whenever its facility identification system procedures flag a difference between the information in the program database and the central registry of shared core facility information. In any event, this issue would be rendered moot under the EIC's recommendation that the notification requirements be changed to allow legal recognition for reporting to a central registry for the overarching data required by all media programs.

[59] See Summary of State Responses, Appendix B.

[60] Ibid.

[61] OMB has indicated to EPA its commitment to the goal of an integrated management system that would improve data quality, provide comprehensive facility information and eliminate duplicative reporting requirements for regulated facilities. In March 2002, OMB directed EPA to improve the utility of the data available on the environmental performance of industrial facilities and urged EPA to develop an integrated system for reporting and access of data across multiple programs Letter dated March 4, 2002, from John D. Graham, Administrator, Office of Information and Regulatory Affairs, OMB, to Kim T. Nelson, Assistant Administrator, Office of Environmental Information, EPA

[62] *Summary of EPA's FY 2005 Budget*, A-3. http://www.epa.gov/ocfopage/budget/2005/2005bib.pdf

[63] *Summary of State Responses*, Appendix B. The State/EPA Information Management Work Group has documented cost savings for states participating in the Exchange Network who eliminate the need for multiple data entries and no longer need multiple interfaces between data systems. See National Environmental Data Exchange Network website. http://www.exchangenetwork.net/benefits/index.htm

[64] One state is currently revising its extensive rules for using and changing its Master File after staff complained that the prior rules have altered and sometimes slowed down the usual business processes within the program areas. By simplifying its rules, however, the agency hopes to address the program staff complaints and to encourage more complete use of the system. Ibid.

[65] Interview with state agency representative (May 2004).

[66] *Summary of State Responses,* Appendix B.

[67] Many centralized state facility ID systems are structured so that the core site facility data is shared among the programs, but is associated with the program interest record that is "owned" by the regulatory programs and can be updated as necessary without changing the fundamental facility ID information. Ibid.

[68] These goals are in keeping with the recent administrative and legislative mandates, such as the President's Management Agenda and the Government Performance and Results Act.

[69] EPA Office of Compliance, *ECHO: Public Comment Period Summary and Response to Comments*, 12.

[70] Examples of facility ID information stored by states in their central data systems include a unique identification number, facility name, alternate name if applicable, facility address, county/municipality, latitude and longitude, short description, affiliations, start date, owner name, operator name, contact name or list, watershed, affiliations, additional ID numbers, type and status (permits, registrations, licenses, legacy system IDs), SIC and NAICS codes,

tribal code, region code, congressional district and site classification.   In EIC's review of state systems, some states with centralized facility ID systems described the process of reaching agreement on common definitions across all media programs as a challenging part of establishing their FIS. *Summary of State Responses,* Appendix B.

[71] Shelley Metzenbaum, "The Evolution of Federal/State Relations—A Win/Win Path Forward," *ECOStates* (Spring 2004), 32 and 35.

[72] AIRNow has been developed by EPA, NOAA, National Park Service, tribal, state, and local agencies to provide the public with easy access to national air quality information. http://www.epa.gov/airnow/partners_list.html

Appendix A

Environmental Information Consortium Membership

and Academy Staff

The Environmental Information Consortium

In 2001, a diverse group of organizations interested in government methods to collect, analyze, and disseminate environmental information held a series of discussions to explore the prospects for "common ground" on recommendations for improving how EPA manages environmental information.   The group included representatives of environmental and public policy groups, the business community, academia, and consulting organizations. They agreed on a broad agenda for reforms and presented them in a letter to the EPA Administrator.[1] A central element of the group's consensus was the need to establish a more comprehensive, unified approach for collecting and integrating facility-specific environmental information in EPA's data systems. The group identified as a priority the need for core identification (ID) data for all facilities regulated by federal environmental statutes to be used by all federal environmental programs.

In response to the group's recommendations, EPA agreed to fund the Environmental Information Consortium (EIC), administered and supported by the National Academy of Public Administration (the Academy). The initial tasks for EIC members were to:

- Identify the needs of all users of environmental data for an integrated facility identification system at the national level, and

- Provide recommendations to EPA on how to establish a feasible and workable system that would meet those needs.

EIC members include those who participated in the initial letter to the Administrator, plus other parties with a strong interest in these issues, especially representatives of state environmental agencies.

Over the past two years, EIC members, with research support by Academy staff, have studied EPA's current facility identification (ID) systems, newer state ID systems, and ongoing efforts to modernize the various databases managed by EPA's media programs. EIC members and Academy staff met with many decision-makers, users, and providers of environmental data, including EPA managers in headquarters and some regions, state agency managers, information

---

[1] In their letter to EPA Administrator Whitman on September 17, 2001, EIC members listed five criteria for operating integrated environmental data system: 1) The data system should be able to provide the environmental information needed by all stakeholders; 2) non-confidential information should be easily accessible and to be easily located, understood and used; 3) users of the environmental data should be able to obtain a comprehensive picture of a facility from one source; 4) the environmental information about a facility should be complete, accurate, timely, and promptly disclosed; and 5) reporting of information by regulated facilities should be efficient, allowing data to be entered once, into a single system. Letter to EPA Administrator Christine Todd Whitman, Administrator, EPA, September 17, 2001.

technology providers, academic or scientific researchers, and representatives from environmental, public interest and business groups.[2]

EIC members asked a number of state agencies questions about the structure of their facility ID systems and how they interface with EPA's data systems.[3] In their responses, state representatives offered many lessons learned in developing, implementing, and maintaining their systems. Throughout this project, the EIC's work has been strongly supported by the Administrator and EPA's Office of Environmental Information.[4]

---

[2] To obtain more candid information, these individuals were assured that their comments would be kept confidential and would not be attributed to specific individuals.

[3] "Summary of State Responses about State Facility Identification Systems, Appendix B.

[4] In a March 2003 letter, EPA Administrator Whitman urged agency cooperation with the EIC project, recognizing that the development of an accurate facility identification system is key to the improved management of environmental information. Memorandum from Christine Todd Whitman, "Support for the Environmental Information Consortium," March 21, 2003.

Environmental Information Consortium Membership

Steve Anderson, Senior Policy Analyst, New Jersey Department of Environmental Protection

Carol Andress, Senior Policy Analyst, Environmental Defense

> Joined Environmental Defense in 1994 as economic specialist working to foster pollution prevention and improve the public's awareness of chemicals in the environment; managed the Great Printers Project, an initiative of local environmental activists, industry groups, state and federal regulators, and labor unions, that reformed reporting to spur pollution reductions and improved the public's understanding of environmental conditions at printing facilities.

Gary Bass, Executive Director, OMB Watch

> Creator of RTK NET (the Right-to-Know Network), a free online computer service to provide community groups with access to government data; former President of the Human Services Information Center; Director of Liaison for the International Year of Disabled Persons; consultant on several projects in special education and the mental health of children and youth.

Dana Bisbee, Partner-in-Charge, Pierce Atwood, Portsmouth Office

> Currently in private practice specializing in environmental and land use law and government relations; served as Assistant Commissioner and Acting Commissioner for the New Hampshire Department of Environmental Services; co-chaired the State-EPA Information Management Workgroup and the EPA/State Stakeholder Forum on Public Information Policies.

John Chelen, Hampshire Institute

> Former Special Counsel and Senior Information Architect for the Office of Technology of the District of Columbia charged with developing a city-wide legal and technical framework for inter-agency information sharing; trained as a systems engineer and lawyer with broad information technology, government affairs and legal experience; led numerous government and commercial system development efforts; developed and managed comprehensive agency-wide data integration and public access systems at DOE, EPA, DOJ, and HSS, and DOD.

Mark Greenwood, Ropes and Gray

> Partner in private law firm Ropes & Gray (Washington DC office) advising clients on a variety of matters related to reporting and recordkeeping as well as public disclosure of environmental information; formerly held a variety of senior positions at EPA in the Office of General Counsel, managing legal issues in areas as diverse as pesticides, toxic chemicals, hazardous waste management, Superfund, and environmental reporting; former Director of EPA's Office of Pollution Prevention and Toxics, concerned with regulating chemicals in commerce, biotechnology, public "right to know" programs and pollution prevention.

Lynn Harris, Chief Information Officer, New Mexico Environment Department

> Director of the Information Technology Division at New Mexico Environment Department; former Chief IT Security and Privacy Officer for the state of New Mexico; over 25 years of IT experience, including work with Los Alamos National Laboratory, the National Center for Genome Resources, and the Department of Education.

Shelley Metzenbaum, Director of the Environmental Compliance Consortium, Visiting Professor and Senior Fellow at the School of Public Affairs, University of Maryland

> Former Executive Director, Performance Measurement Project, JFK School of Government, Harvard University; Associate Administrator for Regional Operations and State/Local Relations, EPA; Under Secretary, Executive Office of Environmental Affairs, and Director, Office of Capital Planning and Budgeting, Commonwealth of Massachusetts.

Betty Miller, Manager, Data Management Unit, New Jersey Department of Environmental Protection

> Ms. Miller has been part of the New Jersey Department of Environmental Protection IT organization for 21 years. She has led many data integration projects including NJDEP's Site Masterfile—the cornerstone for the New Jersey Environmental Management System (NJEMS); participated in several EPA integration efforts; and manages the Data Base Administrators and the Data Quality Assurance Groups.

Melanie Morris, Chief of the Data Integration Division, Mississippi Department of Environmental Quality

> Over 20 years experience in information management and 17 years at MDEQ. Ms. Morris is responsible for all data integration activities and e-government activities within the Office of Pollution Control. Additionally, she participates on numerous national workgroups including the Environmental Council of States (ECOS) Information Management Workgroup and the Network Steering Board for the National Environmental Information Exchange Network.

Sean Moulton, Senior Information Policy Analyst at OMB Watch

> Special expertise in environmental information and right-to-know issues; former Environmental Researcher and Data Manager for the Council on Economic Priorities focused on evaluating and reporting on individual corporate environmental policies and performance; Tax Policy Analyst for Friends of the Earth; Research Fellow and contract employee with EPA's Industry Sector Policy Division.

Paul Orum, Working Group on Community Right to Know

> Senior Advisor to the Working Group on Community Right-to-Know, which defends and improves the public's right-to-know about environmental and public health concerns; served formerly at the Working Group's Coordinator and Director.

Academy Staff

Suellen Terrill Keiner, Vice President for Academy Programs, National Academy of Public Administration

> Former Senior Attorney and Director, Program on Environment, Governance, and Management, Environmental Law Institute; Director of Litigation, the Environmental Policy Institute; Assistant Solicitor and Acting Deputy Assistant Secretary for Energy and Minerals, U.S. Department of Interior; Natural Resources Consultant, Council of State Planning Agencies; Attorney representing environmental and civil rights groups in citizen suits.

Mary Duffy Becker, Senior Consultant to the National Academy of Public Administration

> President, MDB Environmental Law and Policy; former Senior Attorney at the Environmental Law Institute; Associate Attorney, Wald, Harkrader & Ross.

Mark Hertko, Senior Project Analyst, National Academy of Public Administration

> Former Government Relations Intern, Defenders of Wildlife; Quality Assurance/Quality Control Inspector, Accord Enterprises; Community Relations Coordinator Intern, Illinois Environmental Protection Agency; Environment Educator, Illinois EcoWatch.

Appendix B

Chronology of Prior Data Sharing
and Integration Efforts

Chronology of Prior Data Sharing and Integration Efforts

The need for standardizing and integrating environmental data has been discussed for more than a decade. There have been many reports, studies, initiatives, legislative mandates, guidance, and projects on how to improve the organization and sharing of environmental information. Below is a brief description of some of these past integration efforts and projects in which many EIC members have participated.

July 1993: Report of the Task Force of EPA/State Officials to Enhance State Capacity, *Strengthening Environmental Management in the United States.*

> The Task Force looked at ways to improve the partnership between the states and EPA. It recommended that EPA determine what actions needed to be taken to consolidate and integrate its data systems and to facilitate the states' use of those national systems, enabling states and EPA to better share data.

1993 -- 1997: Great Printers Project, a joint initiative of Environmental Defense, the Council of Great Lakes Governors, and the Printing Industries of America.

> The initiative recommended ways to promote pollution prevention and reduce regulatory burden on the printing sector, which is dominated by small businesses. A key outcome of the project was a prototype of a new reporting system, which brought together all of the environmental reporting requirements for the industry in a simplified electronic format. Like TurboTax, the integrated reporting software helped printers to complete regulatory forms while also providing information on pollution prevention techniques that could reduce the environmental impact of their printing operations. Wisconsin Department of Natural Resource provided printers in their state with the software. However, a key obstacle to getting other states to do likewise was the lack of a single identification system for facilities.

August 1994: Report of the National Advisory Council for Environmental Policy and Technology (NACEPT), *Using Information Strategically to Protect Human Health and the Environment: Recommendations for Comprehensive Information Resources Management.*

> The Task Force, which included representatives from industry, states, local governments, the environmental community, and other government agencies, recommended that EPA change its approach to organizing and managing information by establishing an integrated information infrastructure that links information through standardized key data identifiers.

1995 -- 1996: Facility Identification (Key Identifiers) Initiative, *Notice and Request for Comments*, 61 Federal Register 52587 (October 7, 1996).

> Established as part of the March 1995 Reinventing Government II, the initiative focused on having facilities identified the same way for all reporting requirements under environmental laws. The key identifiers approach would have supported one-stop reporting and better public access to information in the data systems, but was not developed by EPA.

March 1995: EPA's One-Stop Reporting Program.

> This program encouraged state-to-state collaboration in addressing environmental information issues and provided funding for states to develop integrated reporting and data management systems.

1995: The Envirofacts Warehouse.

> An integrated database that allowed users to retrieve multi-program environmental information covering federally regulated facilities.

December 1996: A letter from nearly 200 non-governmental organizations was sent to EPA urging strong leadership to reform its facility identification information through integrated key identifiers.

1997: Joint EPA/State group, Facility Identification Template for States (FITS I).

> Through the One Stop Reporting Program, the joint group established working guidelines to help states identify common data elements and proceed with integrating facility site information. The FITS I model developed a data standard based on the Key Identifiers initiative. It has provided important guidance to the states for developing standards and a framework to allow facility-specific information to be shared agency-wide. The template was updated in 2000 (FITS II).

January 1998: Final Report and Recommendations of NACEPT Information Impacts Committee, *Managing Information as a Strategic Resource*.

> The multi-stakeholder group provided recommendations on how to position EPA's information resources for effectively supporting community-based environmental protection and other initiatives, such as the One Stop Reporting Program. The committee recommended creating of an "up-front" national facility registry and restructuring of EPA data systems to allow true data integration. Available at http://www.epa.gov/ocem/pdf/iicfinl.pdf.

1998: Environmental Council of States (ECOS) and EPA jointly formed the State/EPA Information Management Workgroup (IMWG).

IMWG is a group committed to building locally and nationally accessible, cohesive, and coherent environmental information systems. The IMWG created action teams composed of state and EPA members to explore knowledge transfer activities, information exchange technologies, various data systems, and other relevant issues. Available at http://www.epa.gov/OEI/imwg/index.htm.

September 1999: ECOS/EPA, *Environmental Pollutant Reporting Data in EPA's National Systems: Data Collection by State Agencies.*

This joint study supported the need for continued work through partnership to create more efficient and timely sharing of environmental data.

October 1999: EPA created the Office of Environmental Information (OEI).

OEI was established to make better information management a tool for protecting the environment. OEI has worked on many fronts to help facilitate data sharing and to improve the quality, usability, and public access to environmental information. Available at http://www.epa.gov/oei/.

November 1999: Environmental Data Standards Council (EDSC) established.

The State/EPA IMWG formed this council to identify, develop and promote the adoption of data standards that will achieve greater consistency and improve data quality for exchanging information among partner agencies.

Available at http://www.ecos.org/section/projects/?id=1033.

June 2000: Senator Frank Lautenberg, Streamlined Environmental Reporting and Pollution Prevention Act of 2000, Senate Bill 2800.

This bill would require the EPA Administrator to establish an integrated environmental reporting system that would allow a regulated entity to report all required information once each year through one point of contact. The bill also encouraged the use of common data formats and requested EPA to prepare a study on any provisions of law that explicitly prohibit or hinder the integration of reporting requirements. A hearing was held on the bill, but no further action was taken.

July 2000: Exchange Network Blueprint

A team of the IMWG this conceptual design to establish the National Environmental Information Exchange Network. Available at

http://www.epa.gov/OEI/imwg/pdf/final_blueprint.pdf/.

November 2000: National Academy of Public Administration, *Environment.gov: Transforming Environmental Protection for the 21st Century* (Washington, DC)

> The Panel's report urges business leaders, NGO's, and foundations to embrace more effective and efficient policies for environmental protection, and to help build a national data system for gathering, disseminating, and using more accurate and timely environmental information.

January 2001: Government Accountability Office, *Major Management Challenges and Program Risks: Environmental Protection Agency (*GAO-03-112)

> This report cited longstanding information management challenges at EPA.

2001: Facility Registry System.

> EPA introduced this centrally managed database of federally regulated facilities.

April 2001: Environmental Council of the States, *State Environmental Agency Contributions to Enforcement and Compliance*

> This report discussed difficulties in the transfer and accuracy of environmental data.

June 2001: National Academy of Public Administration, *Evaluating Environmental Progress: How EPA and the States Can Improve the Quality of Enforcement and Compliance Information* (Washington, DC)

> The Academy Panel found that both EPA and the states need to invest in better monitoring so they can collect information on actual environmental conditions and compliance with environmental laws.

July to September 2001: Environmental Defense and Tellus Institute convened a group of industry and environmental organizations to promote faster action by EPA on a single identification system. A collaboration of these industry and environmental groups sent a joint letter to Administrator Whitman urging EPA to establish a single facility identification system.

August 2001: The President's Management Agenda was launched as a strategy for improving the management and performance of the Federal Government.

March 2002: Office of Management and Budget, *Letter to EPA*.

> This letter encouraged EPA to improve the utility of available information on the environmental performance of regulated facilities. The letter also urged EPA to establish a single facility identification number and to set up an integrated system for reporting and data access across multiple EPA programs.

2002: Enforcement Compliance History Online (ECHO).

EPA piloted the web-based Enforcement Compliance History Online (ECHO) to provide compliance information on regulated entities.

December 2002: Passage of the E-Government Act (PL 107-347)

This Act mandates greater agency attention to making better investments in information technology investments that will enhance public access and promote delivery of information and issuance of the E-Government Strategy for Expanding Electronic Government.

April 2003: Executive Office of the President, *Implementing the President's Management Agenda for E-Government: E-Government Strategy.*

This report described e-government strategies and challenges across federal agencies.

2003: National Environmental Information Exchange Network.

The first automated data exchange between EPA and states was made through EPA's central node server, the Central Data Exchange.

Appendix C

Summary of Interviews
on State Facility ID Systems

Summary of Responses on State Facility ID Systems

1. Introduction

On behalf of the Environmental Information Consortium (EIC), the National Academy of Public Administration (the Academy) asked managers at state environmental agencies whether they have developed a central data system to identify their regulated facilities. For agencies that have such a system, the EIC asked about their funding and operation. The EIC also sought information on any benefits the state managers have experienced from having the facility ID system (FIS) and any lessons they have learned in its development and operation.

Academy researchers interviewed by e-mail or telephone the Chief Information Officers (or equivalent position) of 16 state environmental agencies, most of whom are members of the State/EPA Information Management Workgroup. To encourage candor in their replies, the state managers were assured that their answers would not be attributed to them unless their state agencies gave specific permission to do so. As a result, this summary of the managers' responses does not identify the states by name. The EIC and the Academy are most grateful to these state managers for their detailed and thoughtful answers.

In their interviews, the 16 states explained how their FISs work and described how their systems endeavor to meet the needs of many different users of environmental data. They also offered a number of recommendations for how to develop and operate a successful integrated system of facility identification information.

The EIC's interview questions were not intended to provide a complete assessment of all state environmental data systems, and the 16 states that answered them do not compose a statistically valid sample. Nevertheless, the 16 responding states do represent a diverse mixture of populations, land areas, geographic conditions, urban and rural economies, and widespread locations throughout the US. Moreover, the state managers' answers have provided very helpful insights for the EIC in preparing this report on how EPA can develop a federal ID system.

Out of the 16 responses, 12 states are currently operating an FIS; and three states are now in the process of developing systems. One state has a system capable of tracking facility-level information across all of its media programs, but the agency is not currently using its system.

2. Development of State FISs

Why develop an FIS?   In most of the 12 states with an FIS, the driving force for establishing a central ID system was to enable agency staff and managers to track all environmental activities at any particular facility and to consolidate data on inspections, compliance status, and enforcement actions across all media programs. As one state manager explained, there is "no good way to integrate activities between programs" unless an agency has an FIS.

These states also mentioned they needed their FISs to improve data quality, reduce data redundancy, and improve overall agency efficiency in managing facility data that had not

previously been shared among their media programs. One state took advantage of revamping its information systems for Y2K compliance to adopt a more integrated data system that included an FIS.

Did you use a vendor? All 12 states used a private vendor in developing and implementing their FISs, and 4 of those states used in-house staff to develop at least portions of their systems.

How long was it in development? On average, development of a state's FIS took 3½ years. For most of the states now operating FISs, initial development began during the period from 1993 to 2001.

Has it been upgraded or adjusted? Since implementing their systems, nine states have made upgrades, adjustments, or enhancements to their FISs to meet specific program needs and to add or improve functions identified through use and testing of the system. One state has made adjustments to the scope of its FIS to respond to legislative mandates for site-specific compliance histories and rankings of facilities.

3. Functions of State FISs

In all the responding states that have an FIS, the system automatically generates and assigns a facility ID number, sequentially or randomly, to a facility when it first enters the state regulatory program or becomes a site of regulatory interest to the state. Facilities may also enter the FIS when their records are electronically "migrated" from various media-specific databases, such as the state air, water, and waste programs, into a central FIS.

Typically, the states' facility ID numbers have no inherent meaning or significance of their own. They are unique to each facility, and the FIS simply uses them to aggregate information about the facility from the media programs' disparate data systems. The FIS attaches or otherwise links its facility ID number to all the media programs' separate identifiers. Some state agencies may also refer to their media programs' data systems as "legacy" data systems or "sub-systems."

How States Initially Populate their FISs

All of the states currently operating FISs initially populated their systems by moving or migrating the basic facility ID information from the various media program databases – "legacy" data systems – into a central FIS. In one state, however, the agency started with legacy data from its water and hazardous waste programs but added brand new data on its facilities regulated for air quality and solid waste.

The state agencies then matched the facility data from their different legacy systems to unique FIS numbers by making manual matches, through matching software, or by a combination of the two. State data stewards or administrators helped to make final decisions about whether possible matches were accurate or whether further clarification about a facility was needed.

Most states agreed that importing facility information from the media-specific legacy data systems into an FIS requires significant effort to clean up the data, maintain consistency, and avoid multiple identifiers being assigned to the same facility. One state employed college interns over a nine-month period to reconcile manually the media programs' files and thus avoided taking program employees away from their regular duties to accomplish the data reconciliation. This effort proved to be more challenging than anticipated, however, because the interns lacked detailed, historic knowledge about facilities that the program employees possessed. This state estimated that 5 to 10 percent of the facilities listed in its FIS are still duplicates, and it will require extensive resources to identify and correct these problems.

Another state used a manual reconciliation process to populate its FIS records by assigning the task to a rotating group of permit writers and compliance staff. These employees reviewed the facilities' individual files in the media-specific systems, reconciled the facility data, and then entered the reconciled data on data-input forms that were used to create a Master File system. The data included all current and historic permits for all media programs at each facility. Reconciling the data to assure its accuracy sometimes required phone calls to the facility managers. Although this process took significantly more time to accomplish than estimated, the agency was not confident in the accuracy of data from any of the existing media programs' databases and thus concluded that manual reconciliation was the best method to assure accuracy.

How States Use and Manage their FIS Records

All 12 states with an FIS use their systems to create a unique identifier for each facility and to link each facility's data among all the media-specific data systems. All 12 FISs hold records containing a core set of general ID information about each facility and provide links to more media-specific information in the underlying program databases. There are some differences, however, in the design of states' FISs and how they manage and use their core facility ID data.

a. Shared Core Facility Data

Six state FISs maintain some type of Master File that enables them to share certain core facility ID data. Their Master Files hold the authorized records of core facility information that are used and shared agency-wide. States have given different names to their records of shared core facility data: a Site Masterfile, Facility Master File (FMF), Core Information, or Client and Site Information.

The basic data elements that most states use to identify facilities generally include site name, location, address, owners, and contacts. All the media programs with an interest in a particular facility or site use the same Master File record for the core data that identify the facility. The media-specific information and different program identification numbers, however, are still maintained by the media programs and are linked agency-wide to the Master File by a facility's unique ID numbers.

As a result, any change to the core ID data on a facility is shared among all media programs and thus alters the basic ID information in their media-specific databases. But other media-specific program records for a facility that do not need to be shared throughout the state agency may be updated and changed as necessary by trained media program staff without changing the core ID records.

b. Warehouse Systems

In six states, the media program offices are not required to share or use the FIS's central facility records. In many of these states, however, the agency has established specific rules that media program staffs must follow whenever they create a new record on a facility that becomes of interest to their offices. Program staffs are trained to check whether the facility is already in the FIS before creating a new facility ID number.

When staff changes existing facility information in the media-specific databases, these states – like EPA's Facility Registry System (FRS) – rely on software that can perform after-the-fact reconciliation. This application automatically flags changes to facility information made by media program staffs and pulls the data into the FIS so they can be checked during a regularly scheduled reconciliation process. During this process, the application will automatically update the states' FIS records or flag the new information for a data steward or administrator to assure its accuracy.

In at least one state, this reconciliation takes place on a nightly basis. Some states also have adopted rules requiring that specific notice be given to the data steward whenever a media program makes any changes to facility ID information.

How States Add a New Facility and Change or Update their FIS Records

Most states with an FIS reported that they have adopted specific operating procedures that all agency staff must follow when adding a new facility or making updates or corrections to existing

facility records in their FISs. Agency staffs must follow these rules whether or not they share the particular ID records agency-wide.

Such state procedures or "business rules" include common language or uniform data standards for entering facility information such as name and address. They also include rules that all program staffs must follow for determining a facility's geographic location or determining how to designate a particular site, such as a mobile drilling vessel.

In most states, only a designated group of trained staff from the media programs or bureau offices is responsible for entering facility identification data in the FISs. Some states even limit this authority to one or two designated staffers, while others authorize several staff from each media program to make these changes.

One agency explained that any staff member who is authorized to update or edit the FIS must use a special security access application to change FIS data, but only a single data administrator can delete a facility from the FIS. Another state that shares core facility ID data allows media program or bureau staff to change facility information on a shared site, but requires that these staff first contact all other bureaus and solicit their comments on the proposed changes before posting any edits in the Master File data.

Some states place strict limits on who can make changes in core ID data for any facility regulated by more than one media program. In one state, if two media programs share data about a single facility, only the data administrator or one "Core Super-User" may change the data. Two states only allow central IT staff to make changes to facility ID data in their FISs. In other states, their software is set up to allow inserts, updates, and deletions to core facility ID data but only if staff follow very specific business rules developed by the agency.

4. Assuring Data Quality

States employ a variety of procedures or checks to ensure data quality for their FISs, which is a significant challenge. They also use these procedures to detect and eliminate duplicate facility ID records. Many states have a data administrator or steward who has overall responsibility for data quality. Two states assign as many as four staffers to monitor additions or changes to facility data and to identify errors in existing program systems.

States rely on specific standard operating procedures for adding, correcting, or updating facility ID records in their FISs, as described above. In addition, many states perform daily quality assurance checks or reviews of any changes to facility data made the prior day. Most state FISs automatically flag any revised data as soon as they are entered, so they can be checked for accuracy.

One state has designated "Site Verifiers" who review and correct data and then mark it as verified. If any data are later changed, the facility's ID record becomes "unverified" until the site verifiers can check it again. Some states report that their FISs maintain log tables that keep records of all old data that have been changed and can be used to track all of the updates made by specific users over time.

States report that their field inspectors or media program staffs most frequently provide updates for facility ID records because they discover changes in basic ID information during their work with particular facilities. In eight states, there is also a regular process for regulated entities to request that the agencies update or correct their ID data. Only two of those eight states allow facilities to request a correction electronically, and none of the states allow facilities to alter their ID data by making entries directly in the FISs.

How States Resolve Discrepancies in Data among Media Programs

When discrepancies occur among facility ID data held by various media programs within an agency, the states take several approaches to resolving these differences. All 11 states that answered this question involve program managers through an internal process for resolving these disputes.

The states' procedures may be either formal or informal, depending on the nature of the discrepancy or the particular state process. One state has adopted a "resolution protocol" that serves as a guideline to resolve these issues in a consistent manner. Another state schedules monthly meetings with representatives from the various media programs to resolve any discrepancies identified during the past month.

Other state answers to this question include:

- "Routine data discrepancies are addressed through the generation of migration error reports which alert legacy system owners of records which require cleanup. New issues require coordination between the program area and the central registry…. Senior management is frequently involved in data quality issues to ensure deadlines are met and issues given appropriate attention."

- "Differing data discrepancies are handled in differing ways. If data discrepancies cannot be handled automatically, they will be handled manually. Facility/site name and address data discrepancies, for example, will be sent to the system's Super Administrator for reconciliation. The Super Administrator will then determine which data is the current "legal name' and address and approve that information. Because of varying facility name issues, the Facility Profiler will keep alias names for user referencing….Major data integrity problems that originate from sub-systems will be passed back through to the sub-system administrators for reconciliation at the sub-system level."

- "If a one bureau data steward identifies a facility shared among bureaus and has data that are inconsistent with data in the database, the data steward is responsible for contacting the other bureaus' data stewards to determine which set of data is most current. When necessary, data stewards contact a facility to verify information. The [state FIS] database has the ability to store alternate and historic data so that bureaus/programs may maintain alternate IDs, names and other information specific to their programs."

- "The programs converse with one another to come to a conclusion regarding discrepancies. When absolutely necessary, IT staff will facilitate this process, often by offering technical solutions."

- "We have no formal discrepancy resolution procedure. When a discrepancy is discovered, IT coordinates between the programs involved to resolve the problem."

How States Resolve Data Discrepancies with EPA

For resolving data discrepancies between the states and EPA, 11 states replied that their approaches are still evolving. Most report that data conflicts can be resolved informally between their state media program staffs and EPA's media programs. In two states, their state data take precedence over EPA data when they encounter such disputes.

5. FIS Facility Data Elements

Most states collect and maintain similar sets of basic facility ID information in their FIS records. Some of these common data elements for identifying facilities include:

ID number
Facility name
Alternate name (if applicable)
Facility address
Locational data (latitude & longitude)
County/municipality
Short description
Affiliations
Start date
Owner name
Operator name
Contact name or list
Watershed
Affiliations
Additional ID numbers, type and status
    (for permits, registrations, licenses,
    legacy system identifiers, etc.)
SIC and NAICS codes
Tribal code
Region code
Congressional district
Site classification

6. Access to FIS Data

Six states let regulated facilities have web-based, 'view-only' access to their own ID data. Although four states do not allow any facility access, two of those four are currently developing web-based access.

7. FIS Funding and Costs

The states' funding for development of their FISs has come from a variety of sources:

- Specific appropriations by state legislatures – 3 states

- EPA grants such as "One Stop" or Network Readiness grants – 3 states

- A combination of federal, state, and program funds – 5 states

- Fees – 1 state

- Penalty funds – 1 state

The states' answers about estimated costs of developing and operating facility ID systems varied widely. The 12 states answering this question spent between $200,000 and $16 million over periods lasting from to two to five years. Of those 12 states, seven states spent more than $3 million.

The states' cost estimates cannot be compared directly, however, because some states have included in their figures all of their costs for modernizing and integrating their entire agency database, not just the cost of establishing an FIS. State costs for migrating facility ID information from their legacy systems into their new FISs also have varied depending on the number of regulated facilities in the state.

In seven states, the staffing costs to operate and maintain their FISs (not including data entry) pay for one to four full-time employees (FTEs), plus contractor support. In any one state agency, the combined annual costs for employing staff and maintaining their FISs have ranged from $100,000 to $750,000.

8. States' Challenges in Developing and Implementing their FISs

The 13 states answering this question encountered similar challenges in the development and operation of their FISs. In seven states, problems with data quality – such as migrating legacy data, duplicate records, and errors in data entry – were a challenge.

For example, one state reported the main challenge has been to recognize that the same facilities had different names in various media program records. This state's IT manager also reported:

Standardized basic information across programs is difficult…[because] some programs such as RCRA require official notification of information changes on their forms before they are willing to change their data. This is particularly obvious when a site changes its name and for a time the different programs within [the agency] will call it different names until they are officially notified of the new name.

In five states, the agencies have encountered some difficulty defining their business processes for using the central FIS and agreeing on common definitions across all the media programs. In addition, one state had to change its business rules for data collection to support an approach focused on using geographic information systems (GIS).

Five states have needed significant cultural changes to convince their staffs to follow the established business processes for their FISs; and it has been important for agency managers to address these cultural issues. As one state explained, one of the biggest challenges was obtaining "on-going commitment by the programs to comply with data standards and making their data available" to the FIS records.

Other state answers about the challenges in establishing an FIS include:

- "The most challenging issues in developing and implementing [Integrated Data for Enforcement Analysis] IDEA has been the migration of data from existing systems, the cost of the contractor who is the only entity able to make changes to the database, lack of appropriate internal project management, and the lack of internal support personnel. This task has been resource intensive (both financial and human). Other challenges have been the lack of general acceptance by staff and mid-level management."

- "Agreement on data standards, cultural challenges around centralized systems and getting the programs to share their data, management commitment, executive sponsorship, data quality, development and commitment to new data entry processes."

- "Many challenges have been met and many lessons learned by all participants. The main challenges are:

  - How to ensure data quality—it is an agency-wide change requiring the development and enforcement of data quality standards

  - How to identify critical data for migration

  - How to resolve cross-organizational issues

  - How to define/document business processes—they must be documented prior to development of an automated process

  - How to resolve conflicting business rules."

9. States' Benefits from their FISs

The states report that their FISs have generated a wide variety of benefits, and they have identified many advantages from using a central facility ID system. The benefit most commonly cited by the states is having easy access to a single ID record that shows all cross-media activities at every regulated facility.

All of the states answering this question report that their FISs enable them to answer information requests more quickly and accurately and with less duplication of effort. In four states, their FISs have improved the quality of their facility ID data. Some states also use their FISs to produce management reports on employee productivity and to track employees' work efforts related to each facility.

Other states report that they use their FISs to:

1) Inform their decisions about setting agency priorities

2) Identify the most serious risks

3) Manage employee workloads

4) Generate information for targeting agency programs: "Reports are generated that indicate which facilities may require attention and where staff resources should be spent"

5) Track fees, fines, penalties, or others paid by facilities

6) Increase agency capacity to analyze facility data, such as adopting GIS applications.

At least two states use their FISs to facilitate and consolidate permitting across all media programs, which eliminates the need for facilities to duplicate their ID information when filing permit applications. One state agency has reduced its permit processing times by 70 percent using consolidated permits that are made possible by its central FIS. Another state reports that its FIS is making possible a merger of records for all public health and environmental facilities into one database.

The 11 states answering this question all stated that they have not tried to estimate direct cost savings or that they find it very difficult to document such savings. Yet, six states report that their FISs have provided indirect savings through increased efficiencies by being able to tie together the regulatory activities in the state across all media.

The states' FISs make staff available for other work because they no longer need to spend many hours manually combining facility ID information drawn from separate databases of the various media programs. This staff time is then redirected more productively to perform other needed agency business, such as analysis of the multi-media data for setting priorities, targeting widespread problems, and identifying serious risks.

One state IT manager reported that, although the agency has not calculated the possible savings from its FIS, "there is cost savings in operating one system for all four major programs rather than maintaining four separate systems, resulting in much duplication of effort." Another state also reported that the more reliable data on facility ID information produced by the central FIS has reduced agency costs for assuring data quality.

Other state answers about the benefits of an FIS include:

- "[The FIS] has enabled agency management and staff to take a broader approach to permitting, field inspections and compliance. Responses to requests for information from the public, the regulated community and from legislators and their staff are accomplished more efficiently. Prior to [the FIS], lengthy file searches were required before inspections just to identify all of the permits held by a particular site. The data was highly dispersed and the data quality varied greatly among the systems. Now, when an inspector wants to find out what permits a regulated entity holds, that information is available in a matter of seconds.

- "[The FIS] is the cornerstone for agency-wide data integration" and "increasing communication across program media lines."

- "[The FIS allows us to] produce the annual report of the department that used to be based on silo-type systems."

- "Allows collection of data only once, easy public and internal access to cross-media information [and] substantially reduced – by 70 percent – the time it takes to process permits."

- "[C]ost savings have not been quantified, but we expect the following efficiencies:

  o "Resource allocation: [FIS] and [agency] on-line resources often replace the need for customers to call and write letters requesting information. As on-line resources become more comprehensive, agency staff resources can be directed away from technical assistance calls, file searches and the preparation of written responses. When agency staff is required to prepare reports, the data is more accessible and takes less effort to extract.

  o Paperflow reduction: [FIS] is a step in the gradual maturation of the [agency] organization from staff-intensive paper flow management to electronic information management. Although paper forms are a component of the process, there is now a *single* facility identifier and a *single* form for changes to core data."

10. Exchanging State Data with EPA

Six states are currently exchanging facility ID information with the FRS at EPA, and five states that do not currently exchange ID data with EPA plan to do so in the future. Four states submit their FIS identifiers to EPA, but they also co-code the records with EPA's identification number or identifiers used by EPA's media programs.

One state no longer submits ID data from its FIS to EPA "because EPA is not interested in our FIS number." Another state proposes to report its core FIS data to EPA just one time and will then use only its own FIS identifier in all subsequent data flows to EPA.

It is especially important to note that, of the eight states answering this question, four states said they have "routinely" encountered poor support and/or technical assistance from EPA when trying to connect their FISs with EPA's data systems. For example, one state said that the agency has problems interacting with EPA data systems due to the complexity and poor documentation of EPA's interface.

Two states have not experienced any problems with EPA's systems, but one state has found that EPA's facility ID data are incomplete. One state noted that the recent change in the contractor operating EPA's central data exchange has seemed to delay the development of services the states are expecting. Another state noted that "EPA does not appear to have adequate resources to support state efforts in moving data via the Exchange Network."

Other typical state responses to this question include:

- "Our difficulties in interacting with U.S. EPA's Exchange Network have involved primarily technical issues in setting up our Network Node and in ascertaining the exact XML formats necessary for data flows. The single facility identifier would not be directly related to resolving those difficulties."

- "The new changeover in staff at the CDX has impacted our sending of data to EPA. Also we were told that the CDX would develop an outbound service, which would allow us to obtain RCRA and TRI from the FRS. They have now indicated that they do not know when this service will be developed and are reluctant to offer an estimate of when it may be available."

11. States' Lessons Learned in Developing and Implementing their FISs

The EIC asked state IT managers what lessons they have learned in setting up their central facility ID systems and what they think will be required to succeed in adopting an integrated facility ID system, either for individual states or nationwide.

The 11 states answering this question suggest multiple ways to ensure the success.
Most states identify the same three key elements for an effective FIS:

- Early and strong financial commitment

- Early and strong commitment and advocacy by senior management

- Support and cooperation from all media programs
.

Many states stress the need for well-documented business rules and processes. They also emphasize that developing these rules requires participation by media program staffs and

flexibility in resolving differences. In four states, the agencies relied on substantial amounts of staff time dedicated to the initial migration of data into their FISs and those dedicated staffs have continued to be critically important for supporting these states' FISs.

Other states' comments highlight the importance of the following FIS features:

- Choosing good technology platforms and ensuring sub-system stability of the data system

- Using technologies that are scalable and flexible

- Setting up robust data sub-systems in each media program

- Planning for continuous improvements, which often cost more than the original FIS

- Paying attention to detail

- Being flexible and adaptable to changes and challenges

- Having patience

One state's experience highlights the importance of high-level leadership and executive support. The state agency is not currently using its FIS even though it is capable of tracking facility-level information across its media programs. This state's manager who answered the EIC's questions reports that agency managers have offered no incentives for staff to enter and track facilities with multi-media interests by using the FIS.

According to this manager, a small amount of work would easily enable the agency to use the FIS for aggregating facility ID information from the various media-specific databases. But this step is unlikely unless the agency's top managers become interested in using the FIS or external influences promote using the system and justify the additional work needed to get it started.

Other state answers to this question include:

- "It requires dedicated maintenance to keep it running properly. The business people don't seem [to] understand how important this data is to keeping their business systems running well or how difficult it is to maintain good quality facility information. 1) Complete management understanding and buy in. 2) Well-documented business process about how the data will be entered and updated. 3) A few staff whose only job is the care and feeding of facility information. A few people need to care deeply about this data."

- "Developing an integrated database requires strong financial and staff resources, and department commitment. It is difficult to work on a project if all resources have not been identified and secured in the initial stages of the project."

- "Support fom management is critical. Be sure to get a complete buy-in from all major program areas. Develop a scope of work that is approved by all. Keep goals simple and achievable."

- "Keep it simple."

- "Keep goals simple and achievable."

- "A good knowledge of the business processes of your agency, executive support that will overcome funding issues and program reluctance, and ultimately program and user support for the process and the product."

- "Get good executive commitment before embarking on the project. Strong project manager with good diplomacy skills, strong program commitment and involvement throughout the project. Budget, Excellent Project Manager who has managed IT projects, Strong Executive Sponsorship and Steering Committee, Program Business area commitment and participation on developing the requirements (Business team), Involve your IT staff from the start as they will have to maintain the system once it is developed."

- "In the course of designing and developing an FIS, iterative review of the system and user feedback is critical as the prototype of the system evolves."

- "Program buy-in, strong management support, a clear definition of the business need and benefits (which will help with program buy-in)."

12. Implications for a National Facility ID System

The EIC asked states about any concerns they might have if they were asked to use a single national facility ID system established by EPA or to adapt their own FISs so they can interact with a national system. Of the 11 states responding to this question, six states had some concerns with a national FIS established by EPA.

First, they said it will be difficult to reconcile state identifiers with EPA's identifiers. They also suggested that there could be difficulty in determining a uniform approach for defining a particular facility and difficult to reach agreement even on what constitutes "a facility." The latter issue includes such questions as whether a very large facility would have one or more identifiers and whether there would be different identifiers for a facility's separate sources of air emissions as well as its other discharge points for water or wastes.   Some states were also concerned about how EPA would deal with the multitude of facilities that only states and not EPA regulate or have an interest in tracking, such as septic systems.

The states also were concerned about whether a national system would increase lag times in the data systems. One state maintained that the states should assign facility IDs because they can respond much more quickly and accurately to changes in a facility's ID data such as ownership, etc., thus minimizing lag times. But they worried that, if EPA is operating a national ID system,

it probably could not make data changes as quickly as the states could. Some states also had concerns about the accuracy of the data if EPA were to manage the ID system.

One state was concerned about the business rules for a national ID system, such as who would have the authority to create, change, or delete a facility ID from the national system. Another state would want EPA to set some data parameters that would ensure that state and federal data systems will be able to interface easily in the future. Yet another state speculated that, if EPA's Facility Identification Template (FITs) model is used to establish the national ID system, there would be few problems regarding data consistency.

Two states were concerned about additional work – although not necessarily extra costs – they might incur to make any changes so their FISs would be compatible with a national facility ID system. One state said the "biggest issue would be internal resistance to the additional work that would be involved with maintaining accurate data across the linked systems."

Additional state responses to this question included:

- "The most significant issue would be reconciling our unique facilities with those recognized by USEPA. We'd also need to deal more quickly with the duplicate facility issue (mostly a resource issue). We probably could adapt our core data to interact with a national ID system; we have very little disagreement over the concept of a facility as presented in FRS."

- "The largest potential issue is that the EPA definition of facility would not be the same as the one we have established in our system. EPA might have two or more facilities where we have one and vice versa. Certainly we could adapt our system to work with a national facility ID system but the costs, confusion and changes to our business process are unknown and are a potential concern. If EPA would mandate this, they should be prepared to assist states with money and or resources to implement it."

- "One can understand that a federal ID may be needed to reconcile facility information across federal systems. [The agency] could adapt its [FIS] to carry these federal ID's but this activity may cause unneeded work. [The agency] will utilize its FIS as the 'master record' for creating a unique State ID and a resource where staff and the public can discover what the [agency] "interests" of that facility/site are. [The agency] handles the State ID/sub-system ID issue by attaching or associating sub-system ID's with the [FIS] State ID.

- "It would be fairly straightforward for us to carry a federal identifier in our database and manage it ourselves, but the states should have the option to assign and manage these identifiers within parameters established by EPA. We have found the ongoing management of identifiers to be fairly complex. Changes of ownership are routine. Facilities split apart and are joined back together. Trying to manage this at the national level would be extraordinarily complex. It's best to manage these at the state level as much as possible."

- One state that is in the planning stages for developing its FIS stated, "Ideally, EPA protocols should have been done before the state develops its own system."

- "That would create issues for us. We are currently creating a single facility identifier for our programs to use. To use an EPA system we would have to reconcile our reconciled facilities with EPA."

13. States without an FIS

Three states answered that they do not currently have central facility ID systems, but all three are currently developing an FIS. One state had already spent 10 months out of a 12-month process to establish a central facility ID system, and another state had issued a request for proposals to develop an FIS. A third state expected its FISs to be operating by spring 2005.

14. Conclusions

The states' answers to the EIC's questions illustrate that there are many commonalities in the development and establishment of an FIS. One very important element is the need to document the different business processes that occur across the media programs within the agency. This documentation offers a good starting point for agencies and their media programs to begin internal discussions about sharing data agency-wide, developing common definitions, and establishing standard operating procedures for using an FIS.

Equally important is the need to have dedicated funding along with strong agency leadership and commitment by key program managers. Implementing new data systems and business processes, plus meeting high staff expectations, require active support by the highest-level agency executives so an FIS can succeed and its many benefits can become a reality. These issues are common among all 12 states that have developed FISs, and they are all the more important if EPA decides to implement a shared national system for facility ID data.

States that have not yet begun or are just starting to develop an FIS will be able to learn valuable information from the above answers to the EIC's questions because other states have already invented the FIS "wheel." Facility ID systems are not simple to build from an information technology perspective, nor are they simple to implement from both a human resource and change management perspective.

There are significant challenges in both respects due to the large amounts of time, money, and staff needed to implement an FIS. Yet, if EPA or states decide to develop a shared facility ID system, they can benefit from the wealth of information that these state managers have already learned for what to do – and what not to do – when implementing an FIS.

Many state managers who answered the EIC's questions have offered to share their experience with others interested in learning from them, and they can provide guidance for EPA and other states to avoid likely pitfalls. Over the long term, a collaborative working relationship among states and between the states and EPA will help to develop the next generation of approaches for integrating environmental data systems and for improving how the states and EPA manage their programs for protecting public health and the environment.

Appendix D

Options for Managing
Facility Environmental Information

Options for Managing Facility Environmental Information

EPA asked the EIC members to consider feasible options for improving EPA's facility ID management that would meet the needs of all users of environmental data. The EIC has identified the following options available to EPA for managing its facility ID data. As stated in its Recommendations section of this report, EIC members have unanimously concluded that Option 3 – a shared Master File System – offers the most efficient, comprehensive, and forward-looking approach to managing and integrating facility ID data among all the media programs.

Option 1: Use FRS to continue back-end data reconciliation and aggregation.

OEI can continue to use the current FRS approach that aggregates and reconciles after-the-fact core facility ID information from the underlying media program databases and then provides public access to that information. FRS is clearly an improvement over EPA's prior attempts at data integration. Despite good efforts and additional resources, however, EPA's current FRS still does not truly integrate environmental information at the facility level.

Rather, FRS relies on back-end or post-collection processes for compiling and reconciling information from the separate media programs' databases. FRS then combines these data only after the media programs have collected them in various forms, and FRS's facility ID number simply serves as a mechanism to link these data. Thus, FRS has not sufficiently reduced the potential for data errors or inconsistencies that lead to inaccurate linkages or information gaps.

Option 2: Increase the frequency of FRS reconciliation and adopt protocols for ensuring changes are made in all the relevant programs' databases.

Another option for EPA is suggested by some of the state systems that the EIC examined during this study. While still using back-end reconciliation, the FRS could reconcile its records more frequently and could adopt protocols or rules requiring that all program staff must notify all other data system managers whenever they create a new facility record or change data in any existing facility record.

Some state systems similar to FRS reconcile their agency-wide records on a nightly basis instead of monthly reconciliations performed by FRS for most of the major databases. More frequent reconciliation would reduce delays between making changes in program databases and making the aggregated data available to the public through EPA's web site. By requiring that program staffs must alert FRS about any changes to facility data changes, EPA could ensure that those changes will be more accurately and rapidly captured in FRS and publicly available on the agency's web site.

While tweaking the current FRS approach could provide some improvement in the quality of FRS records, each EPA media program would still have to duplicate efforts to collect, manage, and update its respective facility ID data. Even if EPA adopts rules requiring staff to share changes in facility data, program staffs would still have no incentives to keep the FRS record updated unless they actually use those data.

Option 3: Establish a single, shared Master File of facility ID information that is the only source of core facility ID data and require that all media programs and regional offices use this Master File.

Rather than having many different "owners" of basic facility ID information, EPA could establish a Master File that would contain the authoritative records of that core data. All media programs would be "users" of the agency's only Master File and would tie their own data to it. This Master File would be similar to some newer state data systems that share their facility ID data agency-wide.   It would also eliminate EPA's current back-end reconciliation process through the FRS. Instead, the Master File system would ensure that all interested parties have timely access to accurate core data for identifying regulated facilities based on uniform definitions and core data elements.

A Master File system would centralize and coordinate the entire process so that errors in facility ID data would be corrected and updates could be verified more quickly and efficiently, rather than relying on each media program to make the appropriate changes. For data reporters, a single, shared Master File would provide a "one-stop" place to record any changes in facility information that would satisfy the legal requirements for all of EPA's offices.