
Facility Identification Template for States
Version 2

***Working Guidelines for Integrating
Facility Identification Information***

Prepared for:

State Environmental Agencies

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

Prepared by:

Ross & Associates Environmental Consulting, Ltd.
under contract to
OAQ Corporation

February 2000

TABLE OF CONTENTS

TABLE OF CONTENTS	I
ACKNOWLEDGEMENTS	IV
GLOSSARY	V
Attribute	v
Data Model	v
Entity.....	v
Relationship	v
Facility Site.....	v
Enterprise Warehouse	v
Hybrid System	v
Central Enterprise Database.....	v
INTRODUCTION	1
WHY AN UPDATE TO FITS	1
WHY INTEGRATE INFORMATION	1
WHY THE GUIDELINES WERE DEVELOPED	2
HOW THE GUIDELINES WERE DEVELOPED AND UPDATED	2
HOW THE GUIDELINES ARE ORGANIZED.....	4
WHO SHOULD READ THE GUIDELINES	4
TERMINOLOGY USED	4
OVERVIEW OF THE FITS II CORE DATA MODEL	5
ORGANIZATION OF THIS SECTION	5
MAJOR UPDATES AND ADDITIONS TO FITS	5
Incorporation of the State EPA Facility Identification Data Standard	5
New Entity Relationship	6
WHAT IS A DATA MODEL?.....	6
HOW TO READ A DATA MODEL	6
Data Model Diagrams.....	7
Attribute Format Tables.....	7
Reference Tables.....	7
CORE DATA MODEL DIAGRAM	8
FACILITY SITE, ENVIRONMENTAL INTEREST, AND ALTERNATIVE IDENTIFICATION	9
Business Need.....	9
Challenges and Recommendations	9
Compatibility with the State EPA Facility Identification Data Standard.....	11
Data Model	12
Entity: Facility Site	13
Entity: Environmental Interest	13
Entity: Alternative Identification.....	14
GEOGRAPHIC LOCATION	14
Business Need.....	14
Challenges and Recommendations	15
Compatibility with the State EPA Facility Identification Data Standard.....	16
Data Model	16
Entity: Geographic Location	17
INDUSTRY CLASSIFICATION	18
Business Need.....	18
Challenges and Recommendations	18
Compatibility with the State/EPA Facility Identification Data Standard.....	20
Data Model	20

Entity: Industry Classification.....	20
AFFILIATION	22
Business Need.....	22
Challenges and Recommendations	22
Compatibility with the State EPA Facility Identification Data Standard.....	23
Data Model	24
Entity: Affiliation.....	24
Entity: Individual	25
Entity: Mailing Address	26
Entity: Organization	26
ALTERNATIVE DATA MODEL APPROACHES	28
FACILITY SITE.....	28
Mobile Sources	28
Search Names	29
GEOGRAPHIC LOCATION	29
Separate Coordinate and Location Address	29
Multiple Geographic Coordinates for a Facility Site	30
FACILITY SITE ASSOCIATIONS	30
FACILITY SITE UNIT.....	31
AFFILIATION	32
Partial Owners of a Facility Site	32
Facility Site-Free Affiliations	32
Organizational Hierarchy.....	33
MISCELLANEOUS ALTERNATIVES	33
Tagging Facility Sites with a Federal Interest.....	33
Merging Geographic Location Addresses with Mailing Addresses.....	33
Tracking Data Timeliness and Quality	33
Providing Staff Contacts for Obtaining Additional Information.....	34
TECHNICAL ARCHITECTURE OPTIONS.....	35
HIGH LEVEL ARCHITECTURE	35
GENERALIZED DATA AND WORKFLOWS	38
Enterprise Warehouse Approach	39
Hybrid Approach	43
Centralized Enterprise System Approach	46
TECHNICAL ARCHITECTURE INVESTMENT DECISIONS	49
Program Interaction and Buy-in.....	49
Funding and Internal Resources.....	49
Consistent Management and Strategic Vision	49
Opportunity.....	49
MANAGEMENT CONSIDERATIONS FOR AGENCY INFORMATION INTEGRATION	51
CRITICAL SUCCESS FACTORS.....	51
Strategic and Tactical Simplicity	51
Leadership and Commitment.....	52
Project Management	53
Communication.....	54
Technology	54
Business Processes and Decision Making.....	55
OVERVIEW OF BUSINESS PROCESS IMPROVEMENTS TO SUPPORT INFORMATION SYSTEMS.....	55
Agency Policies	56
Agency Procedures	58
Program Procedures	61
APPENDIX A: DETAILS OF THE FITS CORE DATA MODEL	62

FACILITY SITE, ENVIRONMENTAL INTEREST, AND ALTERNATIVE IDENTIFICATION	62
Entity: Facility Site	62
Entity: Environmental Interest	63
Entity: Alternative Identification.....	65
GEOGRAPHIC LOCATION	67
Entity: Geographic Location	67
INDUSTRY CLASSIFICATION	72
Entity: Industry Classification.....	72
AFFILIATION	75
Entity: Affiliation	75
Entity: Individual	77
Entity: Mailing Address	78
Entity: Organization	80
APPENDIX B: GENERALIZED ERROR SCENARIOS.....	82
SCENARIO ONE—DUPLICATE CANDIDATES FROM SYSTEM OVERSIGHT.....	82
SCENARIO ONE—ISSUES.....	82
SCENARIO TWO—DUPLICATE CANDIDATES FROM SYSTEM MISUSE	82
SCENARIO TWO—ISSUES.....	83
SCENARIOS ONE AND TWO ANALYSIS—DUPLICATE CANDIDATES.....	83
SCENARIO THREE—LACK OF UPDATE AND MAINTENANCE	84
SCENARIO FOUR—THE RIGHT PAGE	85
APPENDIX C: STATE EPA FACILITY IDENTIFICATION STANDARD.....	86
STATE EPA INFORMATION MANAGEMENT WORKGROUP TRANSMITTAL LETTER	86
APPROVED STATE EPA FACILITY IDENTIFICATION STANDARD.....	89

ACKNOWLEDGEMENTS

The update of these working guidelines has been developed through the support, cooperation, and analytic contributions of a number of state agencies, EPA, and ECOS. The guidelines could not have been developed without the benefit of this participation.

Staff from sixteen state environmental agencies (see list below) contributed to the development of the working guidelines. This Advisory group, lead by Lynn Singleton of Washington, provided documentation and extensive explanation of their states' approaches, offered insight on how to structure the working guidelines to best meet the needs of states, and participated in document review. State participation was critical to this update and their involvement was appreciated. Their input led to significant refinements in the data model and the development of detailed data and workflows.

The State EPA Knowledge Transfer Action Team, co-chaired by Mary McCaffery of EPA and Russ Darr of Washington, sponsored this effort through the State EPA Information Workgroup. The US EPA One Stop Program provided funding through the OAO Corporation for Ross and Associates Environmental Consulting, Ltd. and Windsor Solutions to offer both business process and technical assistance in preparing this document.

FITS II Advisory Group

Dennis Murphy, Delaware Department of Natural Resources and Environmental Conservation
Laurie Beamish, Indiana Department of Environmental Management
Tim Brewster, Louisiana Department of Environmental Quality
Lynn Greene, Michigan Department of Environmental Quality
Dennis Burling, Nebraska Department of Environmental Quality
Pat Hammond, Nebraska Department of Environmental Quality
Betty Miller, New Jersey Department of Environmental Protection
Joanne Jacobs, New York Department of Environmental Conservation
Adele Vogelgesang, Ohio Environmental Protection Agency
Marilyn Best, Oklahoma Department of Environmental Quality
Mitch West, Oregon Department of Environmental Quality
Kim Nelson, Pennsylvania Department of Environmental Protection
Nancie Imler, Pennsylvania Department of Environmental Protection
Gail Jackson, Pennsylvania Department of Environmental Protection
Ken Elliot, Utah Department of Environmental Quality
Troy Delung, Virginia Department of Environmental Quality
Lynn Singleton, Washington Department of Ecology
Keith Borgel, West Virginia Department of Environmental Protection
John Shenot, Wisconsin Department of Natural Resources

GLOSSARY

Attribute

An individual piece of information contained in an **Entity**, e.g., Organization Name, End Date. Properties of **Attributes** include their type—alphanumeric or date—and their length.

Data Model

A model of things of interest to the agency and the relationships between them. The **Data Model** or **Entity Relationship Diagram** uses boxes to represent the things, or **Entities**, and lines to represent the **Relationships** among **Entities**. **Entities** are further broken down into **Attributes**.

Entity

Something that is of interest to the agency, and about which the agency requires information to be recorded, e.g., Facility Site, Organization, Geographical Location.

Relationship

An illustration of how two **Entities** are linked. For example, a Facility Site has one Geographic Location. Properties of **Relationships** include cardinality (i.e., one or more) and optionality (i.e., sometimes or always).

Facility Site

A place where regulatory activities of interest to the agency occur or have occurred in the past.

Enterprise Warehouse

A single enterprise, or agency-wide, warehouse stores a small core set of normalized core data and copies of some program data. Updates entered at the program level are pulled into the warehouse and reconciled. Updates entered at the agency level are not, however, automatically pushed to the program systems. In stead updates from the agency to programs occur through a passive notification and manual update process. Consistency of program systems and the warehouse is enforced by a combination of automated procedures and passive management processes.

Hybrid System

A single enterprise (agency-wide) database master file stores a small core set of normalized data and copies of some program data. The master file is updated from program applications (not databases), from an enterprise data entry application, or both. Updates to program systems are actively pushed from the master file. Consistency of master file data in program systems is enforced by an active, rigorous set of fully automated procedures and business rules, including the ability to override inconsistent changes made to program databases.

Central Enterprise Database

A single enterprise (agency-wide) database is the only authorized store for data, which is used by both agency and program applications. Updates and consistency checks and corrections occur within this single database across program and agency data.

INTRODUCTION

In July 1997, the Environmental Council of States (ECOS), the Washington Department of Ecology, and the US Environmental Protection Agency (EPA) One Stop Program sponsored a workshop in Olympia, Washington. Representatives from twelve One Stop grantee states, five EPA regions, and EPA headquarters attended the workshop and discussed approaches for integrating information about the identity of regulated facilities at sites of environmental interest. In particular, the workshop participants discussed the benefits to be gained by learning from one another's experiences in developing integrated information systems. As a result, the Facility Identification Template for States (FITS) working guidelines were developed in December 1997 as guidance from states to states.

The original FITS document was well received and remains one of the most popular documents downloaded from the ECOS and State EPA Information Management Workgroup websites. Scores of states have documented use of this data model in the development of their own systems or have used the model in requesting proposals from system vendors. EPA's own facility integration efforts have been and continue to be informed by FITS. Further, the State EPA Information Management Action Team used the FITS model and documentation as important building blocks in its deliberations and final approval of a State EPA Facility Identification Data Standard (facility standard).

WHY AN UPDATE TO FITS

FITS was created as a set of working guidelines based on the collective experience of participant states as of mid-1997. Over the past two years states have continued to invest in information management systems and raise the baseline of working facility integration solutions. The update of FITS (FITS II) was prompted by the desire of states to continue to learn from the experiences of states, and to incorporate the data elements and relationships of the facility standard into the template. These are welcome updates and validate the approach of states learning from states, and EPA learning with them to inform its own investments.

This is an update, not a revision; the majority of the FITS model and original technical documentation remain intact. *This update presumes that the reader has read the original FITS document.*

This introduction describes how and why these working guidelines were developed, identifies the intended audience, and describes the overall structure of the document.

WHY INTEGRATE INFORMATION

States and EPA have recognized the need to improve how sites of environmental interest are identified and the information about them recorded. The evolution in environmental management to new strategies, such as place-based and multi-media management, has necessitated that environmental information management also evolve to meet the new decision-making and planning demands. Information systems must enable agencies to uniquely identify sites of interest and relate those to environmental resources and concerns; it is no longer adequate to perceive a Facility Site only through a program-specific lens (e.g., how does the air program define a Facility Site, what are the Facility Site's air regulatory obligations, and how does that compare to the Facility Sites regulated by the waste program). Agencies' interest in improved customer service also necessitates being able to view a Facility Site comprehensively in order to improve and streamline interactions with it. Integrated Facility Site information is a foundation

for a range of new capabilities for environmental management officials, such as consolidated or integrated reporting, targeted outreach, targeted enforcement, and geographically based environmental management approaches.

WHY THE GUIDELINES WERE DEVELOPED

These working guidelines represent a fundamental change in how state environmental agencies traditionally have done business. In the past, agencies typically solved their information problems independently. There is now, however, a growing recognition of the gains to be had from collaborating on information re-engineering and from adopting compatible approaches to information management. Perhaps the greatest gain is the improved ease of sharing information among states and EPA. Another gain is economy: states can adopt proven approaches that not only offer them compatibility but save them the time and money of having to start from the very beginning. A further gain is that troubleshooting becomes easier as the collective experience grows. One's problem is more likely to have been encountered and resolved by someone else, who can then offer guidance.

These guidelines reflect the shift from independence to collaboration in information management. Consistent with this shift, these working guidelines have been designed to present a generic approach to integrating Facility Site identification information. Rather than reflect the specific and detailed needs of one particular state, these guidelines distill the real experiences of a number of states to meet the majority of needs for most states. Individual states may build on the basic approach presented here to meet their own needs.

These guidelines are intended to help states succeed at integrating Facility Site information and, at the same time, help align themselves to be compatible with the State EPA facility standard. However, these guidelines are more comprehensive than the facility standard, which identifies a minimum set of data elements that would be desirable for states and EPA to share about sites of environmental interest. See Appendix C for more details. Developed by states for states, these guidelines identify the types of information that states likely would find desirable to maintain for their sites of environmental interest. Their objective is a workable approach that will enable states to better identify, understand, and interact with their Facility Sites.

States that have not already done so are advised to begin integrating their Facility Site information. Proven state approaches are available, several of which are distilled in these guidelines. The longer states wait to begin, the more non-integrated, non-standardized, program-specific data they will have to tackle in the transition.

HOW THE GUIDELINES WERE DEVELOPED AND UPDATED

The original guidelines were developed through discussions with and a review of documentation from the states of Massachusetts, New Jersey, Oklahoma, Pennsylvania, and Washington, as well as EPA. On the basis of that review, the core data model was developed using the following criteria:

- Only implemented, well-proven approaches drawn from states were considered for inclusion in the core data model. This criterion reflects the belief that the core data model should not recommend that other states pursue an untested approach. Nonetheless, unproven and alternative data model approaches are presented in the "Alternative Data Model Approaches" section of the document.
- The data model addresses only basic information needed to identify and interact with a Facility Site. The "80 percent" rule has been applied, such that the core data model focuses

on basic Facility Site identification information that most, if not all, states likely would want to maintain in order to successfully identify and interact with a Facility Site. States with more elaborate information requirements can build on the core data model.

- Where a type of Facility Site information can be handled in multiple ways, the core data model presents a preferred approach, selected on the basis of states' experience transitioning, maintaining, and more importantly, using the data in question.
- The core data model is lean and pragmatic, designed to accommodate the majority of Facility Sites in the environmental universe. Inevitably, a few Facility Sites will not fit easily. The simplicity of the model, however, makes it much easier for users to understand and apply successfully and consistently.

The original FITS document was released in 1997. In 1999, the Facility Identification Action Team completed the facility standard and forwarded it to the State EPA Information Management Workgroup for approval. Under the guidance of Lynn Singleton of Washington, a proposal was put forward to the Workgroup's Knowledge Transfer Action Team that the original FITS states, along with other interested states, revisit and update the FITS document to reflect the facility standard and the states' collective experience over the past two years. As with the original FITS document, the EPA One Stop Program supported this work. The Knowledge Transfer Action Team approved a state advisory group consisting of Delaware, Connecticut, Indiana, Louisiana, Michigan, Nebraska, New Hampshire, New Jersey, New York, Ohio, Oklahoma, Oregon, Pennsylvania, Texas, Utah, Virginia, Washington, West Virginia, and Wisconsin. The advisory group shaped the basic approach to the update effort and periodically reviewed the working drafts. Staff were directed to do the following:

- Complete a review of the original FITS document.
- Analyze elements and definitions of the newly approved facility standard, including a review of the content area expert discussions used in its development and refinement.
- Structure interviews with the advisory group states and others. For states with significant production experience with Facility Site data, the interviews focused on lessons learned over the past two years. For states still early in their efforts, the interviews focused on key questions and interest areas that an updated FITS should address.
- Conduct informal discussions with One Stop (and other) states, especially those who had requested technical assistance from the One Stop Program, to assist them in refining an approach to integration of Facility Site data.

The information collected in these tasks was used to identify and develop the specific updates considered by the advisory group

The update process was guided by the original FITS philosophy of focusing on demonstrated solutions in use by states, stressing simplicity, and presenting preferred and alternative options. Unlike the discussion of the core data model, the new discussion of architectural approaches does not specify a preferred option. As discussed below, architectures turn out to be dependent on many factors not relevant to the template itself.

In addition, comments were solicited at the Washington State Facility/Site Workshop. The advisory group completed their deliberations and review of the draft in late January 2000 and the final document was released in February 2000.

HOW THE GUIDELINES ARE ORGANIZED

The guidelines are organized into the following major chapters:

- Overview of the FITS II Core Data Model—the overall considerations that influence the structure of the core data model, differences between the core data model and the standard, specific data encompassed by the model
- Alternative Data Model Approaches—other unproven approaches to consider
- Technical Architecture Options—three basic technical approaches, with associated data and work flows
- Management Considerations for Agency Information Integration—internal policies and procedures that agencies should consider to help ensure the success of their integrated systems
- Appendices—additional information for the consideration of readers
 - Appendix A-- detailed definitions and formats for the entities and attributes that make up the core data model.
 - Appendix B-- generalized scenarios highlighting common system errors
 - Appendix C-- the State EPA Facility Identification Data Standard

WHO SHOULD READ THE GUIDELINES

The policy discussions and technical details in these guidelines are intended for managers, program staff, and information resources staff of state environmental agencies. Both policy decision-makers and information technology experts should review these guidelines and jointly help the agency consider their applicability.

Much of the text is intended to help the reader consider how best the state can succeed with integration from a policy and organizational perspective. Technical details are provided in appendices, which may be of particular interest to information resources staff.

TERMINOLOGY USED

Although discussions with states and EPA have used the term “facility,” the working guidelines have adopted the term “Facility Site,” which implies more of a sense of place, regardless of any physical structure at that place. The term Facility Site, as used in these working guidelines, is intended to capture what media programs might variously call a facility, a source, a point, etc. As defined in the Glossary, a Facility Site is “a place where regulatory activities of interest to the agency occur or have occurred in the past.”

OVERVIEW OF THE FITS II CORE DATA MODEL

The working guidelines set forth the FITS II core data model. It is a template that states can use for defining, relating, and integrating information on the identity of their Facility Sites. This chapter introduces the concept of a data model, describes the associated business needs, and summarizes its content. Each section includes a discussion about key issues of the FITS II data model and the facility standard and how a state implementing a Facility Site system can ensure interoperability. These discussions identify and expand the facility standard as a basis of interchange and define its relationship to the critical business needs of most agencies. Detailed technical information about the core data model is provided in Appendix A.

ORGANIZATION OF THIS SECTION

After defining a data model and explaining how to read one, this chapter identifies and describes the major components of the FITS II core data model:

- Facility Site—the essence of what the programs consider a “facility,” a “source,” a “site,” etc.
- Environmental Interest—program-specific interests in a Facility Site (e.g., does the Facility Site have an air permit, is the Facility Site a cleanup site)
- Alternative Identification—identification numbers already in use by the various programs (e.g., in permit files), as a “pointer” towards more detailed program-specific information about the Facility Site
- Geographic Location—physical location of the Facility Site
- Industry Classification—type of industrial activity conducted or services produced at a Facility Site
- Affiliation—roles and responsibilities of external organizations and individuals associated with a Facility Site (e.g., owner, fee contact).

The information about each of these data model components is provided in a standardized way:

- Business Needs—state environmental agency need
- Challenges and Recommendations—issues associated with maintaining the information and FITS II’s handling of those issues
- Compatibility with the State/EPA Facility Identification Data Standard—correspondence of the core model to the facility standard and effects on the implementation of FITS II
- Entities and Attributes—types of information represented by the data model component.

MAJOR UPDATES AND ADDITIONS TO FITS

The original FITS data model, now with two years of scrutiny from practical state application, remains a credible and evolving template. As part of that evolution, several elements have been either updated or added for clarity, including the incorporation of the facility standard and new relationships between data model entities.

Incorporation of the State EPA Facility Identification Data Standard

The FITS document was used extensively by EPA in its efforts to develop a data standard and by the State EPA Facility Identification Standard Action Team. Many of the features of the facility standard were derived from or based on the FITS model. However, it is important to recognize a critical if subtle distinction between FITS and any data standard. FITS is a template for states to consider in designing systems to *store and manage* facility data. The facility standard is best thought of as a template for the *interchange* of facility data between two parties. FITS II

presumes that an interchange via the facility standard is an important business need for states. The following changes, based on the facility standard, were made to the original FITS data elements to assure compatibility with the standard:

- “Format” changes (in attribute type or length)
- Addition of several attributes in the facility standard but not in the original FITS core model

However, attributes in the original FITS core model but not in the facility standard remain in the FITS II core model. In addition, attributes that exist as single elements in the facility standard but are captured as multiple elements in the FITS model remain multiple elements in the FITS II core model (e.g., Address Line 1 and Address Line 2 rather than Address only).

These changes minimize any sacrifices to the simplicity of model itself while ensuring that states using FITS II guidance, will, with minimal effort, be able to interchange data, based on the facility standard.

New Entity Relationship

Over the past two years, states have both confirmed the original FITS entity relationships and developed new ones to fit their own dynamic business needs. FITS II contains one new multi-option relationship for Industry Classifications. As with the original treatment of Affiliation, the FITS II model now supports the entity-Industry Classification to have a relationship with both Facility Site and Environmental Interest. This allows agencies to manage both affiliations and industry classifications as single “enterprise” assertions and/or separate the environmental interest-linked assertions of these entities. These relationships are emphasized not because these “triangles” should necessarily be implemented in any state’s physical data model, but rather these assertions are central to the FITS approach itself. States should decide how to manage the information and how the model, application logic, and business process will be supported. All other entity relationships were found to be consistent with the original FITS data model. Other relationships in use by one or more states were reviewed and added to the “Alternative Data Model Approaches” section of the document.

WHAT IS A DATA MODEL?

A data model is a generalized plan of the information that a database system should contain and the interrelationship of that information within the system. The data model is made up of entities, attributes, and their relationships. An entity is a general identifier of a concept—in FITS II, the entities include Facility Site, Environmental Interest, and Geographic Location among others. Attributes are the specific pieces of data that describe each entity.

The FITS II core data model draws on the experiences of twelve states to provide a generic representation of the entities and attributes that states will want to consider for inclusion in a site identification database.

HOW TO READ A DATA MODEL

A data model has three major pieces: the data model diagram, attribute format tables, and reference tables. These are explained below. The following descriptions are provided for the benefit of the lay reader and should not be interpreted as “how-to” guides for systems development. The interpretation and use of a data model for systems development is best left to highly trained individuals.

Data Model Diagrams

A data model diagram, which is formally known as an Entity-Relationship Diagram (ERD), is a way of illustrating the types of data to be contained within a system and how these data elements are interrelated. ERDs follow strict formatting conventions so that the contents and relationships are always understood consistently.

An ERD has three parts: boxes, lines, and “crow’s feet.” Each box in an ERD represents an entity. A line between two boxes represents a relationship between the entities contained in the boxes. A simple line between two boxes reflects a one-to-one (1:1) relationship: entity X can be related to only one entity Y, and vice versa (for example, an American can only have one Social Security Number). Sometimes, however, a line will have “crow’s feet” on it, which illustrates a one-to-many relationship (for example, an American can have many cars). The crow’s feet attach to the “many” entity. In the figure below, the boxes represent the Facility Site and Environmental Interest entities in FITS II. The line between them shows that these entities are related. The crow’s foot at the end of the line next to the Environmental Interest box means there can be multiple Environmental Interests for a single Facility Site.

Attribute Format Tables

The Attribute Format Table defines the format of each attribute. The table contains three columns (length, type, and requirement) which together define the attribute’s format. Length indicates the maximum number of numerals or characters contained in an attribute. (For example, the attribute for Mailing Address State Name has a length of 35 characters). Type indicates the nature of the data value. “Alphanumeric” allows any combination of characters. “Date” allows only valid dates. [Length is not defined for any FITS II Date attributes because length typically is predefined by the individual database.] Requirement indicates whether the attribute referenced must contain data. If an attribute’s requirement is considered “optional,” absence of data is not critical to understanding the context of the entity information. If an attribute is considered “mandatory,” the entity information is essentially useless without the attribute having been populated (for example, knowing that an individual exists is only meaningful if the name of that individual also is known—name therefore is “mandatory”).¹

Reference Tables

Reference tables provide the permitted values for a given attribute. Attribute format tables are found in Appendix A, “Details of the FITS II Core Data Model.”

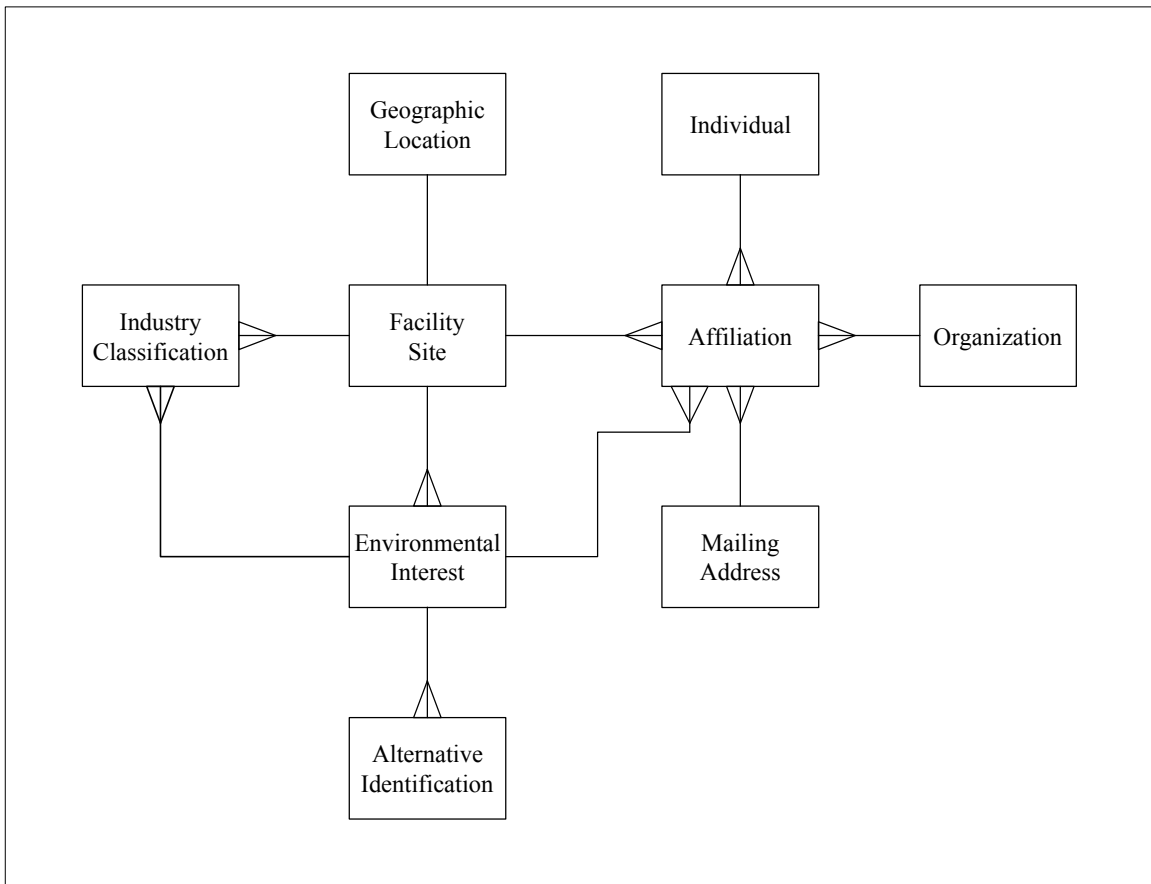
The FITS II core data model does not attempt to provide permitted values for attributes, since these vary according to an agency’s business needs. Nonetheless, states are strongly encouraged to rely on existing standards wherever possible, both to simplify system development and to enhance compatibility for data sharing with other states and EPA. For example, to automate county information, a state could rely on established FIPS county codes rather than creating a new county coding system that would require translation to FIPS for any data interchange. Moreover, as more states and EPA begin to integrate site information, it will be valuable if states and EPA can work together to identify some core set of permitted values in key areas of the site identification information. This will simplify data interchange, help stakeholders more readily understand information across states and nationally, and eliminate needless repetition and duplication.

¹ The term “requirement” does not refer to a requirement, imposed by EPA or by state legislation or regulation, that the state collect this information. The term rather reflects an assessment of whether the information is essential or optional to having viable information about the entities in the data model. This is described further in “Details of the FITS II Core Data Model.”

For definitions of some of the attributes below, the FITS core data model includes the statement: “the permitted values for this attribute, along with a description of each, should be recorded in a separate look-up table. This suggests a requirement to keep such descriptive data (e.g., description for each Standard Industrial Classification Code) in a part of the database other than that which would support the entities and attributes specified below. These look-up tables have not been included in these guidelines, for two reasons. First, to include them would require the addition of many other entities that would drastically reduce the readability of the document without adding much substance. Second, the way that such information is modeled for a state’s system can vary depending on its own technical standards (e.g., each look-up table could be modeled in the system architecture as a separate entity, or they could be consolidated into a “generic code” entity). As a general rule, codes in information systems are useful as a space-saver and as shorthand for more sophisticated users. However, coding is not very helpful for many users, especially for an integrated system intended to serve a broad audience. Regardless of how the state addresses the need for look-up tables, it is essential that descriptive fields be provided in the system for the range of users who will find shorthand codes meaningless.

CORE DATA MODEL DIAGRAM

Simplified ERD



FACILITY SITE, ENVIRONMENTAL INTEREST, AND ALTERNATIVE IDENTIFICATION

Facility Site, Environmental Interest, and Alternative Identification are discussed together because of their relationship to what a Facility Site is and an agency’s interest in it.

Business Need

The term Facility Site is intended to capture the combined essence of what programs refer to as a “facility,” a “source,” a “permitted site,” an “operation,” a “cleanup site,” etc. An agency cares about Facility Sites that have had, are having, or could have an impact upon the environment due to activities that have occurred or may occur there. An Environmental Interest is a need or activity that causes a program to be interested in a specific Facility Site; it is usually related to a regulation or regulatory program that applies to a Facility Site (e.g., Superfund site, Title V permit). Environmental Interest information provides the agency with the ability to track details (e.g., contacts, permits, status) about a Facility Site that are only relevant because of an interaction by the specific Environmental Interest.

Challenges and Recommendations

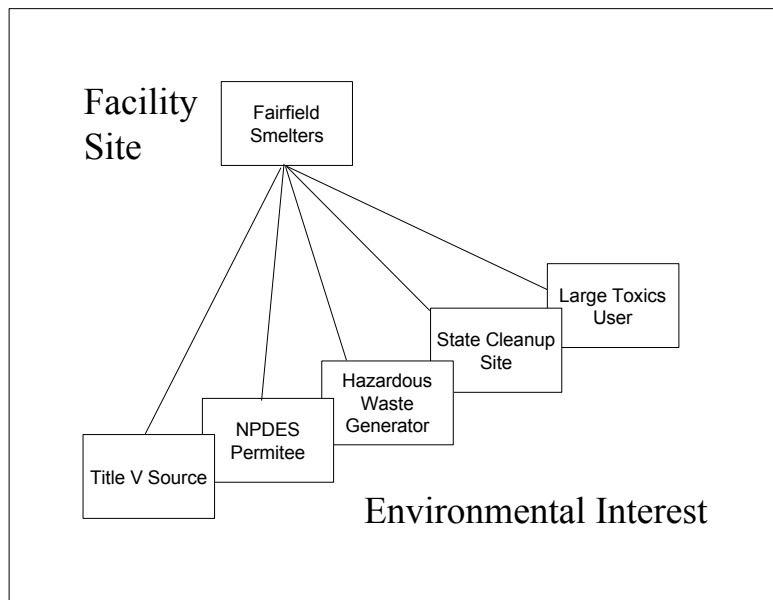
Where is a Facility Site

A Facility Site is generally considered to have a fixed location. The boundaries of the Facility Site, however, may not be easily discerned or even known by the agency. Each program may have a different perspective of what bounds a Facility Site and those boundaries can change over time. FITS II addresses the location of a Facility Site by providing a geographic location. This can be represented in up to three ways: as a geographic coordinate, as a physical address and description, and/or as a geopolitical location. The “Alternative Approaches” section discusses, in greater detail, options for how states could accommodate mobile sources.

Eliminating Duplicate Facility Sites

Currently, many programs may collect identification information regarding an individual Facility Site. Much of this information, such as an address, may be duplicated and difficult to cross-reference across programs.

Facility Site – Environmental Interest Relationship



States typically want a system that refers to each Facility Site by a unique identifier—one that references only the site in question and no other. Using that unique identifier, the public can access all of the environmental conditions and activities at Facility Sites, regardless of which program has regulatory authority. For further discussion about the process for uniquely identifying Facility Sites see the “Technical Architecture Options” section.

Historical Facility Sites

A Facility Site may be of a specific interest to the agency for a finite period of time. For example, a Facility Site might be an active cleanup site for 18 months or until the agency deems that it has returned the spill site to accepted ambient conditions. The Facility Site may become another cleanup site later, or may involve concurrent cleanups. Agencies typically wish to track such information. The system must, therefore, accommodate the status and related dates of a Facility Site’s interest to the agency. FITS II provides for status dates associated with each Environmental Interest and a date qualifier to aid the interpretation of those not familiar with the details of the Environmental Interest data

The date provision does not necessarily support such instances as when a Facility Site changes its name but not its interaction with the agency. To accommodate the full spectrum of potential Facility Site name changes, an additional attribute is required in the data model. This is accomplished through a listing similar to an alternative ID or through a direct attribute field for historic names. Other states simply track dates of operation under a specific name. Generally, program areas associated with Superfund or other cleanup type activities have a direct business need to populate the attribute.

Naming Facility Sites

Programs may find that the name used for a Facility Site is different from that of other programs. Furthermore, while the programs may agree with the intent to have a consolidated version of the Facility Site’s name as part of the integrated identification data, they may need time to adjust to a new “agency” name for the Facility Site.

FITS II allows for alternative names, called Facility Site Alternative Names. This capability allows programs or other users to query the system by the name most familiar to them.

Facility Site Identification Number(s)

The concept of having a single consolidated view of a Facility Site leads to the need for a unique identification number for a Facility Site. This provides the reference number that anyone can use to communicate an interest in a specific Facility Site. Whereas a Facility Site’s name can sometimes be ambiguous (many different Facility Sites can have very similar names), a Facility Site’s unique identification number cannot.

The FITS II core data model provides a state unique identification number for each Facility Site.² The transition to reliance on unique Facility Site identifier can require a significant amount of time and is best achieved incrementally. During the period of transition, support must be provided to accommodate both the new unique Facility Site identifier and the existing alternative

² EPA is planning to assign its own unique identifier to sites in the foreseeable future. This is to reassure EPA of stability in its own systems during the transition to integrated data from the legacy systems, during which time states and EPA will be working out the data interchange requirements and data stewardship procedures for state-EPA data sharing of integrated site information. The FITS II core data model has accommodated a federal identification attribute, called Site Federal Identifier, in the event that a state wants to keep track of the federal ID number in the state system as a way of obviating any potential confusion in state-EPA data interchange during the transition.

methods of identification. Typically, any given Facility Site carries multiple existing program-specific identification numbers: the number(s) by which it is known in state program systems, in federal systems, in permits, and in paper files. As states integrate their Facility Site data, they need to make links between the integrated Facility Site data and the legacy or older program-specific data held in various state and federal program systems. These links serves two purposes: they help users of the integrated Facility Site information locate more detailed information about the Facility Site's Environmental Interests and they help the agency's programs recognize the now-integrated Facility Sites. States that have implemented a unique number have expressed a caveat to not to publicly release the unique identifier until the agency is ready to begin using the number for some central receiving function.

The data model recommends that agencies accommodate multiple Alternative Identification numbers for each Environmental Interest in a Facility Site. For example, if a Facility Site is of interest because it is a Large Quantity Generator of hazardous waste, it may have four identification numbers: one each for the RCRIS and TRIS federal systems, another for the state billing system, and yet another for the state regional office. Because these numbers in and of themselves may be meaningless to the data user, the data model should provide for an explanation of the purpose or origin of a given identification number. For example, the number "PAD001296729" is a RCRA identification number, for which detailed information can be found in the RCRIS system.

Environmental Information System Identifiers

The FITS model provides significant flexibility in how the various legacy identifiers for a Facility Site and Environmental Interest are linked. However, it is essential to FITS that each Environmental Interest be uniquely identified for each Facility Site, and that the definition be consistent.

Many program legacy systems never distinguish between Facility Site and Environmental Interest-- every Facility Site had a permit and if it had two, it was in the database twice. In most cases, the permit number, facility number, and environmental interest number are the same attribute. In more complex situations, however, there may be one generic Environmental Interest that involves multiple sub-interests or the permit number for a given Environmental Interest may change over time as the permit is re-issued. In these cases, states have used a separate ID for the Environmental Interest Identifier (Environmental Information System Identification Number), which can be linked (either in the Alternative Identification section or in the program system itself) to these program specific identifiers. The state may also choose to link these more specific Environmental Interest identifiers to Site Units (e.g. the individual tanks at a Facility Site) as indicated in the alternative approaches section.

Compatibility with the State EPA Facility Identification Data Standard

Facility Site

There are two minor differences between the facility standard and the FITS II model. Physical address information is a part of the Facility Site group in the standards, but is not a part of the Facility Site component in FITS II. In the FITS II core model, physical address is captured in the Geographic Location component.

The other difference is that the standard's "Facility Registry Identifier" attribute makes a specific reference to EPA's Facility Registry System (FRS). Because FRS has not been implemented, it is not used in the FITS II core model. The equivalent attribute in the FITS II core model, "Facility Site Federal Identifier," instead refers to a generic Federal ID number.

In both cases simple transformations allow data stored in the FITS II template to be made compatible with the standard. Use of the facility standard would entail merging the physical address information from the FITS II Geographic Location component into the facility standard's Facility Site group. In addition, the FITS II Facility Site Federal Identifier attribute would be converted to the appropriate Facility Registry Identifier number in the facility standard as soon as EPA's FRS is implemented.

Environmental Interest

The following attributes have been added to the FITS II template to ensure consistency with the facility standard: Environmental Interest Start and End Date Qualifier and Environmental Information System Abbreviated Name. These attributes qualify dates to aid in the interpretation of detailed Environmental Interest data. Depending on the architecture chosen by a state, these attributes may or may not be a part of that state's core models. However, this information can often be obtained through more detailed information stored in specific program systems (some states may have to reconstruct this information). States beginning to design and develop new systems should consider including these elements as part of the core set of information.

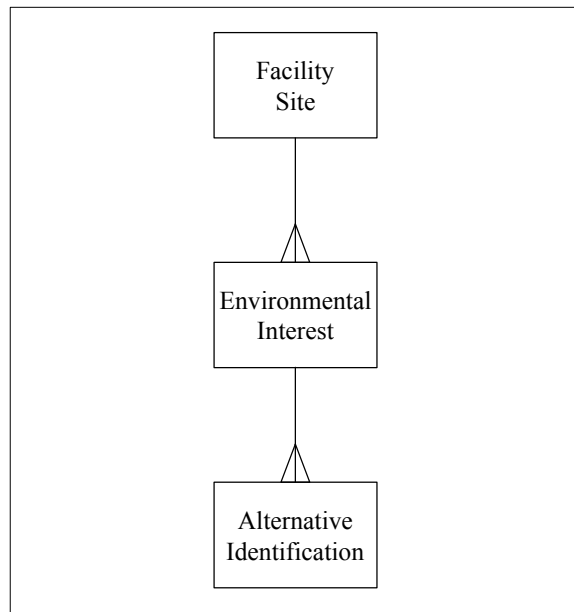
The standard does not support the expression of an Environmental Interest associated with anything other than a Facility Site. By contrast, the FITS II core model allows for Environmental Interest to be associated with Facility Sites and with Affiliation. For interchange, the Affiliation-Environmental Interest association information is dropped or translated into an Environmental Interest-Facility Site association.

Alternative Identification

The standard does not have a component equivalent to the FITS II core model's Alternative Identification component.

Data Model

Data Model Diagram



Entity Definitions

Entity: *Facility Site*

Definition: A place where regulatory activities of interest to the agency occur, or have occurred in the past.

Usage: This definition is intentionally vague. Although the vast majority of Facility Sites are easily identified (e.g., a factory), some are harder to discern. This is partly due to the complex structures of some Facility Sites (e.g., non-contiguous land, multiple buildings) and to the varying and sometimes conflicting definitions specified in environmental rules. Experience has shown that, even for the more complex cases, Facility Site identification can be achieved, through a case-by-case approach, if necessary.

Relationships between Entities:

- A Facility Site is always monitored because of one or more Environmental Interests
- A Facility Site always is found at one Geographic Location
- A Facility Site sometimes performs activities associated with one or more Industry Classifications
- A Facility Site sometimes involves one or more Affiliations

Attributes:

- State Facility Site Identifier*³
- Facility Site Federal Identifier⁴
- Facility Site Name*

Entity: *Environmental Interest*

Definition: An Environmental Interest is a need or activity that causes a program to be interested in a specific facility site; it is usually related to a regulation or regulatory program that applies to a Facility Site (e.g., Superfund site, Title V permit).

Usage: Environmental Interest provides the agency with the ability to track details (e.g., contacts, permits, status) about a Facility Site that are only relevant due to the interaction provoked by the Environmental Interest.

Relationships between Entities:

- An Environmental Interest always represents an interest at one Facility Site
- An Environmental Interest sometimes requires a program to interact with one or more Affiliations
- An Environmental Interest sometimes has one or more Alternative Identifications
- An Environmental Interest sometimes tracks one or more Industry Classifications for a specific Facility Site

Attributes:

- Environmental Information System Identification Number*
- Environmental Interest Facility Site Alternative Name
- Environmental Interest Type*
- Environmental Interest Start Date*

³ * indicates that the attribute exists in the facility standard

⁴ Equivalent to the "Facility Registry Identifier" attribute in the facility standard.

- Environmental Interest End Date*
- Environmental Interest Start Date Qualifier*
- Environmental Interest End Date Qualifier*
- Environmental Information System Abbreviated Name*

Entity: ***Alternative Identification***

Definition: An alternative identifier (in addition to the Environmental Information System Identification Number) by which a particular program that has an interest in the Facility Site knows of it. The identifier may be a permit number issued by the program, an identification number for the Facility Site used in the program's files or database, a historic Facility Site name, or a Federal database identification number (or any other identifier a state finds useful to include).

Usage: Many of these alternative identification numbers may be needed to support links to 'legacy systems.' Over time, these forms of identification may become unnecessary as agency-wide use of the State Facility Identifier becomes standardized.

Relationships between Entities:

- An Alternative Identification is always an alias used by a program with an Environmental Interest in the Facility Site

Attributes:

- Alternative ID Number
- Current System Code
- Facility Site Historic Name

GEOGRAPHIC LOCATION

Business Need

Agencies depend on Facility Site location information in many ways:

- To search for a Facility Site by its location (e.g., the Facility Site nearest the stream or aquifer)
- To describe the location of a known Facility Site (e.g., Where is Acme Plating?)
- To confirm the identity of a Facility Site (e.g., Is this the Facility Site located at 110 Elm St.?).

These needs are distinct from those met by a Facility Site's Mailing Address. Geographic Location is used to *locate* a Facility Site, while the Mailing Address is used to *correspond* with individuals and organizations affiliated with a Facility Site.

The Geographic Location of a Facility Site can contain three kinds of related locational information:

- Physical address and location description
- Geographic coordinates and their associated Method Accuracy Description (MAD) Codes
- Geopolitical areas such as congressional and legislative districts.

Challenges and Recommendations

Collection and management of a Facility Site's geographic location data presents several challenges and opportunities. Specific programs maintain locational data of varying format, quality, and coverage. For example, some programs may contain good latitude/longitude (or other coordinate system) data for many Facility Sites, while other programs may contain only unformatted address information. In developing a single set of Facility Site locational data, agencies need the ability to reconcile among various formats. Address matching applications can automatically convert address and coordinate data. This provides a means both of filling in one when the other is missing, or when reconciling data from multiple programs, matching and confirming the identity and location of a shared Facility Site. Address and coordinate data can be used to derive most geopolitical areas. Given a Facility Site's coordinates, one can easily compute the Facility Site's corresponding congressional and legislative districts.

Within the physical address data, states are likely to find widely varying formats, especially for the more detailed street and route level information. A relatively generic structure for capturing this data (the Address Line 1 and Address Line 2 attributes recommended here) provides the greatest flexibility to capture such variances. The growing sophistication of address matching applications allows all but the most unusual addresses to be parsed automatically from such attributes. The core data model also recommends that states record only a single reference point coordinate for a Facility Site. A more complex model which supports the storage of lines and areas is described in the "Alternative Data Model Approaches" section.

Migration into the new agency-wide integrated data system of program geographic coordinate data (where it is available) is also likely to produce data of various formats. In addition to (or instead of) latitude/longitude coordinates, many programs and states use UTM, Range/Section/Township, and Quarter codes. While the core data model stores only the decimal latitude/longitude, it is recommended that the state Facility Site system be able to accept, translate, and display multiple coordinate system formats. This is important both to facilitate the migration of program data into the system and to ensure that the programs can view the data in the format they require.

MAD codes are metadata (i.e., data that describes data) that store information about the method, accuracy, and description of geographic coordinates. Many possible MAD codes can be associated with any given Facility Site's geographic coordinates, depending on how the various programs created the data in their own systems. Given the great number of possible MAD codes, the data model identifies a minimal set that experience suggests are most likely to be used and correctly supported by states.⁵ In many cases, even those programs fortunate enough to have significant locational coverage often lack metadata about these coordinates. In such cases, the MAD codes provide a place to say what is known about the quality of the data (i.e., not much) and to record new, better-qualified information as it is collected. In many cases, the MAD codes will be highly interrelated; for example, common collection methods (e.g., the use of a given GPS unit or address matching application) may have default associated accuracy, units, and source scale codes that can be calculated automatically.

As metadata, MAD codes are often neglected. They tend to be much more important to other data users than to those who originally enter the data. As a result, making the entry and maintenance of MAD codes as painless as possible is especially important to ensure the overall utility of the data.

⁵ For discussion about additional MAD codes please see the "alternative approaches" section.

Geographic location data is so crucial that the data model recommends it be a mandatory data element. In other words, a Facility Site record cannot be created without a corresponding geographic location record with the address, location description, or coordinate attributes provided.

Compatibility with the State EPA Facility Identification Data Standard

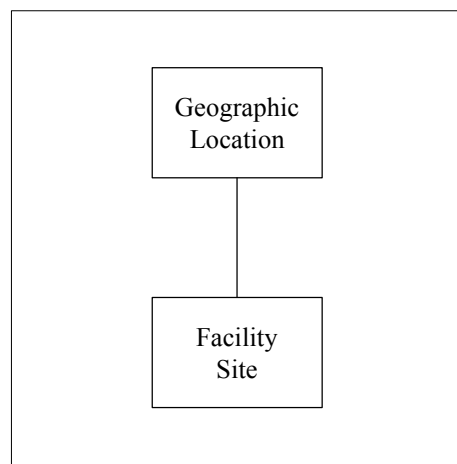
The standard and the FITS II core data model apply a distinctive organization to a common set of Geographic Location information, both in terms of what information is stored and the relationship of that information to other components. The standard defines the Geographic Coordinates group as a geographic point, or set of points, defined by latitude and longitude coordinates used to locate a Facility Site, including the associated method, accuracy, and description data. The standard does not include physical address as part of the Geographic Coordinates group (this information is stored in the Facility Site group.) The FITS II core model allows each Facility Site to be associated with only one Geographic Location, defined by three kinds of related locational information (the physical address, location description, and geographic coordinates; MAD codes; and geopolitical areas). For further discussion about the options and reasons for separating coordinate and location address information as well as allowing multiple geographic coordinates for a single Facility Site, please see the “Alternative Data Model Approaches” section.

The following elements have been added to the FITS II template for consistency with the facility standard: Country Name, Tribal Land Name, Geometric Type Code, Geometric Type Name, and Reference Point Source Map Scale Number.

Translation from a FITS II template into a format compatible with the facility standard could be accomplished by merging the physical address information from the Geographic Location component with the Facility Site information to provide the equivalent of the facility standard’s Facility Site group. Since FITS II recommends only one Geographic Location for each Facility Site, there are no associated translation issues. Conversely data states receive in the standard format may contain multiple locations. This data would have to run this data through their normal reconciliation and resolutions process to select a single location for each facility site.

Data Model

Data Model Diagram



Entity Definitions

Entity: *Geographic Location*

Definition: The physical location of a Facility Site, which may be described using up to three types of information: physical address and description, geographic coordinates, and geopolitical areas.

Usage: Provides the physical location of the Facility Site. Attributes of Geographic Location are highly interrelated.

Relationships between Entities:

- A geographic location is always associated with one Facility Site

Attributes:

- Location Identifier
- Location Address Line 1*
- Location Address Line 2*⁶
- Locality Name*
- Location State USPS Code
- State Name*
- Location ZIP Code/International Postal Code*
- Location ZIP Code +4 Extension
- Location FIPS County Code*
- County Name*
- Country Name*
- Tribal Land Name*
- Supplemental Location Text*
- Latitude Measure*
- Longitude Measure*
- Geometric Type Code*
- Geometric Type Name*
- Horizontal Collection Method Code*
- Horizontal Collection Method Text Name*
- Horizontal Accuracy Measure*
- Horizontal Accuracy Measure Units
- Horizontal Reference Datum Code*
- Horizontal Reference Datum Name*
- Reference Point Code*
- Reference Point Text*
- Reference Point Geographic Position Description
- Source Map Scale Number*
- Congressional District Number
- State Legislative District Number⁷

⁶ “Location Address Line 1 and Line 2” attributes are equivalent to the State EPA Facility Identification Data Standard’s “Location Address” attribute.

⁷ Each state may have different legislative-type systems. ‘State Legislative District Number’ should be viewed as a generalization of how state political information is collected. Actual data elements should be configured to fit the business needs of the agency.

INDUSTRY CLASSIFICATION

Business Need

Industry Classification refers to the nature of the industrial activity performed at a Facility Site. Agencies have a need to understand the Industry Classification for various reasons. First, numerous regulatory requirements are industry-specific. Most agencies have a business need to understand which regulations apply and impact a Facility Site. Second, industrial processes that generate environmental impacts typically are common to an industry. By analyzing Facility Sites across an industry, agencies can obtain a better understanding of behavioral norms and variances and can design and target appropriate regulatory responses.

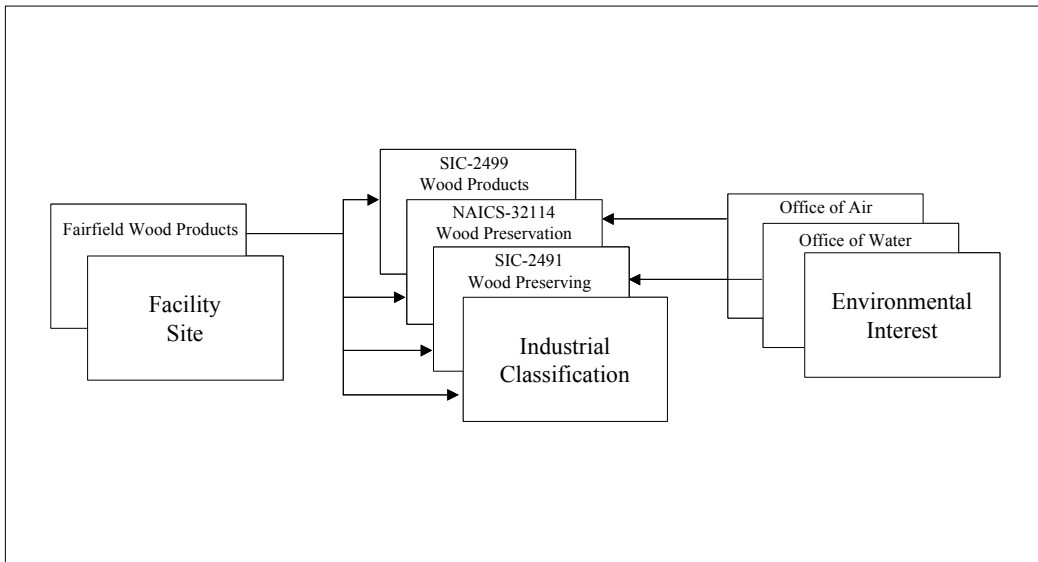
Environmental agencies traditionally have characterized industrial activity through the use of Standard Industrial Classification (SIC) codes. Facility Sites are required to report the nature of their industrial activities using one or more SIC codes. The North American Industrial Classification System (NAICS) code (intended to eventually replace the SIC code) is beginning to be implemented in some federal government agencies. Several states are planning migration efforts from SIC to NAICS over the next several years.

Challenges and Recommendations

In transitioning to integrated Facility Site information, agencies will be drawing on program-specific data. Many programs are likely to hold one or more Industry Classifications (potentially SIC or NAICS codes) for a given Facility Site, gathered through reporting associated with a specific program interest. Some programs will have no Industry Classification data. In integrating the program-specific data, agencies will need to make decisions about how many Industry Classifications to maintain for each Environmental Interest and/or Facility Site.

The following figure illustrates that Industry Classifications can be associated directly with a Facility Site, Environmental Interest, or both.

Industry Classification Relationship



Industry Classification could be captured in a myriad of ways through the Fits II Model. In the example above, the Facility Site, Fairfield Wood Products, is associated with three different Industry Classifications (two SIC codes and one NAICS code). Fairfield Wood Products classifies itself as primarily a wood products company. In most of the interactions with the state agency (e.g., general permits), this is the recognized industrial classification (SIC-2499). However, the Office of Water classifies Fairfield Wood Products as a wood preserving company (SIC-2491) because the business treats its products with various chemical preservatives. The Office of Water has a permitting relationship with the company because of these chemicals. In addition, the Office of Air began using NAICS codes and classifies Fairfield Wood Products using the NAICS code for wood preservation (NAICS-321114).

To effectively maintain the integrity of industry relationships, a set of business rules would need to be defined. These rules need to specify how enterprise determinations of Industry Classification are made and how many Industry Classifications can be maintained for each Environmental Interest. There are several additional issues that the business rules will need to address: 1) Will SIC, NAICS, or both be tracked by the agency as an enterprise core element; and 2) If both types of codes are to be stored will a link between the two be maintained.

States will need to evaluate when and how to best transition from SIC to NAICS codes. Cross-reference tables are available to assist in the conversion of SIC to NAICS codes. The experience of several States suggests that these tables do not provide an exact translation mechanism and that some manual intervention will be needed.⁸

This data model allows agencies to use and maintain multiple Industry Classifications for both Environmental Interests and Facility Sites. This is a significant change from the original FITS model, which recommended associating industry classification information ONLY with Facility Sites. Any Industry Classification relationships should be driven by a specific agency business need. To effectively manage this information, separate tables for SIC and NAICS codes and appropriate cross-reference tables will be needed. Because there is no one-to-one relationship between SIC and NAICS codes, it is not recommended that the information be stored in the same table. The number of Industry Classifications to be tracked for each Facility Site-Environmental Interest is also a matter of agency preference. If the agency is using SIC codes, they should be at least two digits long (the least specific format), but ideally four digits in length for greater utility. Some states use up to a six digit SIC code. If the agency is using NAICS codes, the correct six-digit code should be identified. The data model suggests that a Facility Site be assigned multiple Industry Classifications to portray the range of its industrial activities. The model also suggests identifying a primary classification for a given Facility Site. "Unknown" may be used as a valid entry whenever the primary industrial activity is not easily determined. Such usage should support the analytic needs of the agency.

Finally, states should consider likely future enterprise data flows they may wish to accommodate. Many states are considering (or have piloted) new, integrated reports or consolidated service centers. These data sources may represent new ways agencies collect and use enterprise information like Affiliations and Industry Classifications on behalf of the agency.

⁸<http://www.epa.gov/idmdssc1/html/sicsum.htm> Additional information about the NAICS Codes and the SIC Codes can be found at this web site, which contains the document "Summary Report of North American Industrial Classification System and Standard Industrial Classification Data Standards." It contains listings of standard data elements for the NAICS and SIC codes, a data model of the relationship between the different codes, and linkages to other web sites with information on the SIC and NAICS codes.

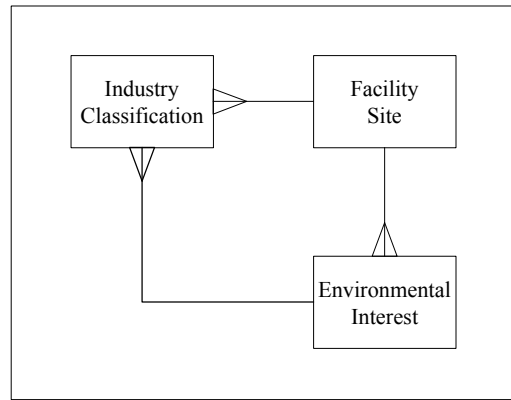
Compatibility with the State/EPA Facility Identification Data Standard

The facility standard treats SIC and NAICS as two separate groups, while the FITS II core model discusses both in the Industry Classification component. This is a subtle difference, however, the FITS II model does not recommend that SIC and NAICS information be stored and managed in the same location in an implemented Facility Site system. Ideally, there would be separate tables for SIC and NAICS codes and the appropriate cross-reference tables linking the information to Facility Site and/or Environmental Interest. SIC and NAICS codes are discussed together, in the FITS II component, because of the similar nature of the information represented.

The standard does not support the expression of an Industry Classification associated with any entity other than a Facility Site. The FITS II core model allows for associations to Facility Sites, Environmental Interests, or both. The Industry Classification-Environmental Interest association information would be dropped or translated into the Industry Classification-for the Facility Site during data interchange.

Data Model

Data Model Diagram



Entity Definitions

Entity:* *Industry Classification

Definition: The type of industry or business conducted at the Facility Site.

Usage: This provides a general description of the type of activities conducted at the Facility Site.

SIC and NAICS codes are used to describe the type of industry. Because various activities may be performed at a Facility Site, more than one SIC or NAICS code may apply to any one Facility Site. Industry Classification may also be associated with specific Environmental Interests.

Relationships between Entities:

- An Industry Classification sometimes describes the type of activities performed at one Facility Site
- An Industry Classification is sometimes associated with a specific Environmental Interest

Attribute:

- Standard Industrial Classification (SIC) Code*
- SIC Primary Indicator*

- U.S. National Industry Classification System (NAICS) Code*
- NAICS Primary Indicator*

AFFILIATION

Business Need

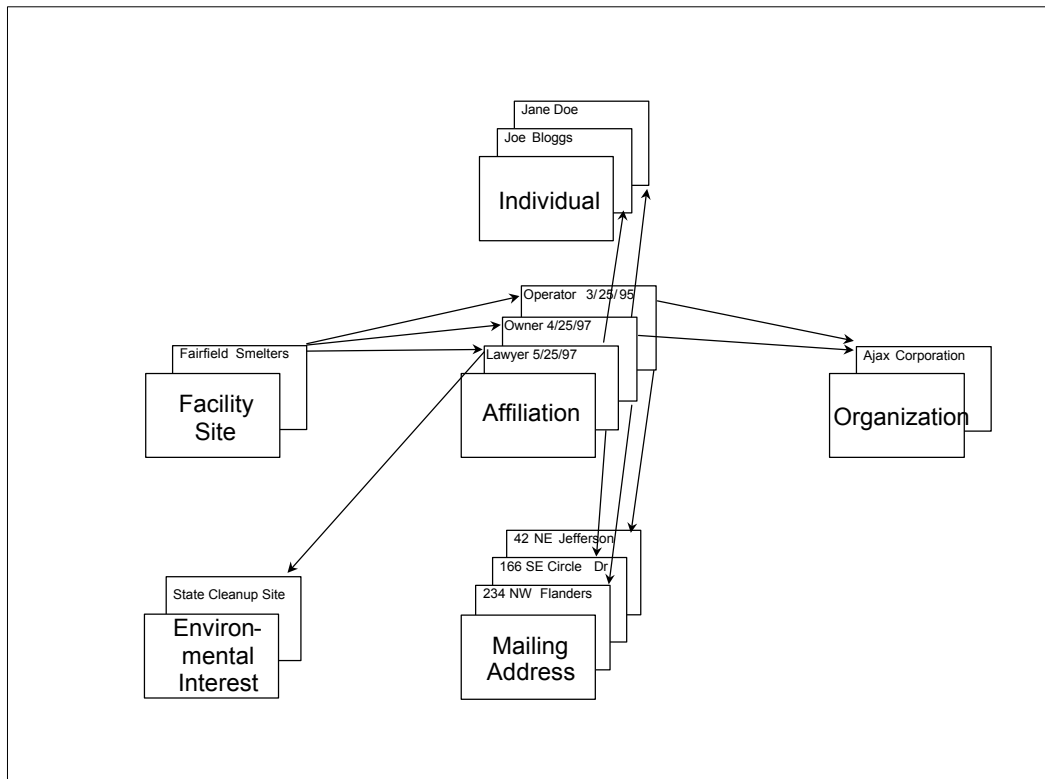
Agency programs do not correspond with Facility Sites, they correspond with Individuals and Organizations affiliated with the Facility Sites. Agencies typically need to know many types of information about such affiliations, such as: names of the parties responsible for regulatory compliance (e.g., the business owner, the operator), and the contact people (e.g., for billing, for hazardous waste information)⁹.

Challenges and Recommendations

Many Affiliations may be associated with one Facility Site. In order to contact these affiliates, the agency must have either a mailing address or a telephone number.

The following figure illustrates the concept of Affiliation. The figure shows how Individual, Organization, Mailing Address, and Affiliation are related to a single Facility Site.

Affiliation Relationships



An Individual or Organization may act as the sole affiliate to the Facility Site (e.g., owner). Even when an Organization is the affiliate, a specific Individual may be included as the point of contact for that Organization. (For example, the owner of the Fairfield Smelters Facility Site is the company Ajax Corporation; when the agency wants to contact that owner, it is useful to know

⁹ Note that ‘interested parties’ such as public interest groups are not considered part of this information. They are excluded from the scope of this data model due to the different nature and quality of their information. This data model only captures information about individuals directly related to the Sites.

that Jane Doe can represent the company for such interactions.) Therefore, any one Affiliation may include an Individual, an Organization, or both.

Although an Organization (and sometimes an Individual) may have multiple offices, each with a different mailing address, only one mailing address needs to be recorded for each Affiliation. An Organization may provide a different mailing address to the agency for communication regarding a *different* Facility Site. Furthermore, an Organization may also provide a different mailing address for communication regarding its performance of a *different role* for the *same* Facility Site. For example, whereas the Ajax Corporation wants correspondence regarding its ownership of the Fairfield Smelters Facility Site to be mailed to its Flanders Street office, it wants correspondence regarding its role as the Operator of that same Facility Site to be sent to its Jefferson Street office.

For some Facility Sites, a separate Affiliation(s) may need to be recorded for each Environmental Interest. When consolidating this information across programs, individual programs generally collect the same basic identification information for a Facility Site. However, in a large Organization, one person is not always responsible for the entire environmental management of that company and therefore could not serve as the contact for all of the agency's varied interests at the Facility Site. The company may have one person responsible for hazardous waste, one for air pollution, etc., which requires the agency to record a separate Affiliation for each Environmental Interest at the Facility Site. For the cleanup being performed at Fairfield Smelters, the Lawyer is Joe Bloggs; however, Joe Bloggs may not represent Fairfield Smelters for any of its other environmental activities. FITS II supports multiple Affiliations related to the Facility Site overall or to individual Environmental Interests.

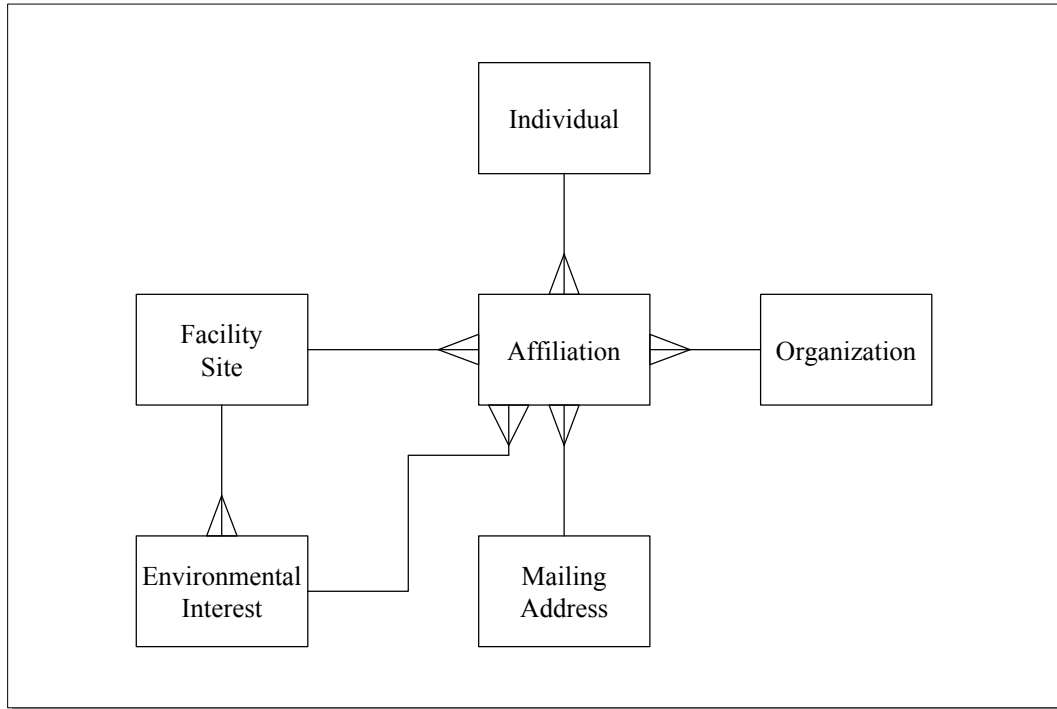
Over time, Facility Site ownership changes, and people leave companies. In some such cases (e.g., for the responsible party of a cleanup site), the agency is required to keep track of the history of changes. Therefore, for each Affiliation, the agency may record the start and end date of the Affiliation.

Compatibility with the State EPA Facility Identification Data Standard

The standard contains a subset of the Affiliation attributes found in the FITS II core data model. Also, the FITS II core model supports Affiliations associated with both Facility Site and Environmental Interest. The standard allows Affiliation to be associated with only Facility Site. This Affiliation-Environmental Interest association information is dropped or translated into an Affiliation-Facility Site association, during data interchange.

Data Model

Data Model Diagram



Entity Definitions

Entity: *Affiliation*

Definition: A relationship between a Facility Site and an Individual or Organization.

Usage: The Affiliation defines the roles that external Organizations and Individuals play at a Facility Site. Such relationships between Organizations, Individuals, and a Facility Site provide the agency with the names, phone numbers, and mailing addresses that the agency needs in order to communicate with a Facility Site. Some Affiliations represent current relationships with the people affiliated with the Facility Site, and some represent past relationships.

Although each Affiliation specifies only one Organization or Individual, the same Organization or Individual may be associated many times with the same Facility Site through other Affiliations. This feature accommodates situations where the same Organization or Individual plays many roles at a Facility Site (e.g., owner, operator, and site contact). Furthermore, any one Organization or Individual may be affiliated with many Facility Sites. This is particularly important as it allows the agency to determine all the roles an Organization or Individual plays with Facility Sites in the state.

The Affiliation Type (see attribute below) will likely dictate how an Affiliation should be recorded given the agency’s rules for programmatic sharing of, and historical tracking of, Affiliation. More specifically, the Affiliation Type will

determine whether an Affiliation is one shared by the entire agency or is specific to one environmental interest. The legal owner of a Facility Site should only need to be recorded once, regardless of each Environmental Interest in the Facility Site, whereas the Facility Site contact may be different for each environmental interest in the Facility Site, and need to be recorded several times.

In addition, Affiliation Type will determine whether the agency must track the history of previous affiliations. When the responsible party for a Facility Site no longer plays that role, the agency must record that the Affiliation has ended and the date on which it ended, but cannot remove any reference to it. However, when a Facility Site Contact has left the site, the agency does not need to track the contact's historical Affiliation with the Facility Site.

These rules may be enforced by formal working procedures or may be enforced by the system. If they are enforced by the system, a separate look-up table is needed for each Affiliation Type.

Relationships between Entities:

- An Affiliation always involves one Facility Site
- An Affiliation sometimes involves an Individual
- An Affiliation sometimes involves an Organization
- An Affiliation sometimes involves a Mailing Address
- An Affiliation is sometimes associated with a specific Environmental Interest

Attributes:

- Affiliation Type*
- Phone Number
- Phone Number Extension
- E-mail Address
- Fax Number
- Alternate Phone Number
- Alternate Phone Number Extension
- Affiliation Start Date
- Affiliation End Date

Entity: Individual

Definition: A person that is of interest to the agency because of that person's interactions with one or more Facility Sites (e.g., a person who is the contact for three Facility Sites and is the responsible party for another).

Usage: It is important for the agency to avoid duplication of Individuals, so that the agency can answer questions such as, "What other sites is this Individual involved with?"

Relationships between Entities:

- An Individual is always a person involved in one or more Affiliations.

Attributes:

- Individual Identifier
- Last Name*
- First Name*

- Middle Initial*¹⁰
- Individual Title Text*

Entity: Mailing Address

Definition: The standard address used to send mail to an Individual or Organization through the US Postal Service.

Usage: A mailing address is primarily used as the means for the agency to formally correspond with an affiliated Individual or Organization. A Mailing Address is not intended to allow the agency to physically locate an affiliate (or even a Facility Site), but only to allow them to mail something to the affiliate.

For international mailing addresses (which typically represent only a very small percentage of cases), the country must be specified. Where the format of the address does not match the attributes below (e.g., instead of city, a district or other jurisdiction type), such information should be included in the Address Line 1 and Address Line 2 attribute fields. This would include mailing codes for international addresses—ZIP codes are only required for US addresses.

Relationships between Entities:

- A Mailing Address is always the address for one or more Affiliation

Attributes:

- Mailing Address Identifier
- Mailing Address Line 1*
- Mailing Address Line 2*¹¹
- Supplemental Address Text*
- Mailing Address City Name*
- Mailing Address State USPS Code
- Mailing Address State Name*
- Mailing Address ZIP Code/International Postal Code*
- Mailing Address ZIP Code +4 Extension
- Mailing Address Country Code
- Mailing Address Country Name*

Entity: Organization

Definition: A company, government body, or similar entity that has some responsibility or role at one or more Facility Sites (e.g., an Organization that operates five Facility Sites and is the owner of two of them).

Usage: It is important for the agency to avoid duplication of Organizations, so that the agency can answer questions such as, “What other sites is this Organization involved with?”

Relationships between Entities:

- An Organization is always involved in one or more Affiliations.

Attributes:

- Organization Identifier

¹⁰ “Individual Last Name, First Name, and Middle Initial” attributes are equivalent to the State EPA Facility Identification Data Standard’s “Individual Full Name” attribute.

¹¹ “Mailing Address Line 1 and Line 2” attributes are equivalent to the State EPA Facility Identification Data Standard’s Mailing Address attribute.

- Organization DUNS Number*
- Organization Formal Name*

ALTERNATIVE DATA MODEL APPROACHES

For states' consideration, this chapter of the working guidelines identifies alternative approaches to integrating Facility Site identification information. As previously explained, the core data model reflects only those integration and data model approaches considered both proven and generic on the basis of states' experience implementing them. These alternative approaches are presented as either unproven or optional alternatives that are of interest to some states.

Because most of the alternatives described below are additions to the FITS II core data model, they may increase the complexity and overhead of managing the state system. Given that states are generally best served by performing an incremental transition towards integrating their Facility Site information, they should defer the implementation of any alternative unless their need for it is clear and well understood. The lessons a state will learn about such needs once they have access to even a basic integrated Facility Site identification system will provide important insights into how the system should be expanded to accommodate more elaborate needs.

The alternative approaches are organized in the following main categories:

- Facility Site
- Geographic Location
- Facility Site Associations
- Facility Site Unit
- Affiliations
- Miscellaneous Alternatives

FACILITY SITE

Mobile Sources

The modeling of Facility Site does not implicitly accommodate mobile sources (e.g., rock crushers, asphalt plants). However, regulatory requirements such as licensing, may generate a state need to record mobile sources.

Several state approaches have emerged as alternatives. Most states have developed a flag on Environmental Interest Type that would indicate whether the source is mobile or portable. That flag could allow the Facility Site to exist in the database without an associated Geographic Location. Some states flag the Facility Site and list the last known location or the location of the owner or operator. A few states actively track the portable source from Facility Site to Facility Site. States that have chosen this latter approach have found it difficult to maintain updated information.

One state allows Environmental Interests to be associated with MULTIPLE Facility Sites (each with begin and end dates). As an object of interest moves (e.g., a mobile incinerator) it is linked to the appropriate Facility Sites (if they exist already) or to new "virtual" Facility Sites if they do not. This allows the source to be tracked and easily identifies those places where the mobile source is collocated with a traditional stationary Facility Site (i.e. a contaminated site).

Search Names

Agency information users may wish to have multiple versions of a Facility Site's name to support for more flexible searching of the database (e.g., including a full text search name when a Facility Site's name includes an acronym). Some of these capabilities (e.g., phonetic searches) are subject to the technology in use at the agency and to the user interface standards that the agency employs. A policy on search may raise issues about agency versus program-specific data, (in this case, names) and how much the agency is willing to accommodate. One risk associated with this type of functionality is that it may not be well maintained, and after a number of years could include many inappropriate search names.

GEOGRAPHIC LOCATION

States have taken a wide variety of approaches in organizing the geographic location information about Facility Sites. These variations are often related to the state's use of location data by other non-Facility Site related data (e.g., monitoring stations), the state's history of collecting location data, or the state's use of GIS.

Separate Coordinate and Location Address

While both the latitude/longitude coordinates and the location address of a Facility Site are related (one may even have been derived from the other), they are different kinds of data. States may wish to reflect this difference by separating the latitude/longitude coordinates and MAD code data from the location address information. Separating coordinates from location address could support further uses of coordinate data, such as if the state wants to address Facility Site Units (refer to the discussion below) in its data model. The Facility Site Unit entity could be related to the Geographic Location entity so that units could have their own coordinates but not their own location address. To separate coordinates from location address, the Geographic Location attributes presented in the FITS II core data model would be separated as follows:

Entity: Geographic Coordinates

- Latitude
- Longitude Measure
- Geometric Type Code
- Geometric Type Name
- Horizontal Collection Method Code
- Horizontal Collection Method Name
- Horizontal Accuracy Measure
- Horizontal Accuracy Measure Units
- Horizontal Reference Datum Code
- Horizontal Reference Datum Name
- Reference Point Code
- Reference Point Text
- Reference Point Geographic Position Description

Entity: Location Address

- Location Identifier
- Location Address Line 1
- Location Address Line 2
- Locality Name
- Location State USPS Code
- Location ZIP Code/International Postal Code
- Location ZIP Code +4 Extension
- Location FIPS County Code

- Supplemental Location Text
- Location FIPS County Code
- Congressional District Number
- State Legislative District Number

Multiple Geographic Coordinates for a Facility Site

The separation of Geographic Coordinates may allow the model to associate multiple coordinate points at a single Facility Site. This would commonly be used to represent lines and boundaries of a Facility Site. While the data models of some states support this feature, its use is not widespread, largely because most states do not have data to support multiple coordinates. Further, several states have noted that this level of complexity may be better handled in GIS rather than in a tabular data system as presented in the FITS II core data model.

The facility standard allows for multiple Geographic Coordinates to be associated with a Facility Site. However, each instance of a Geographic Coordinate can only be associated with one Facility Site occurrence. If multiple coordinates are associated with a single Facility Site, the Geometric Type Code and Name attributes would be used to classify the geometric entity (i.e. point, line, area, region, or route) represented by the sequence of coordinates.

FACILITY SITE ASSOCIATIONS

Facility Site Associations are a grouping of Facility Sites, the interrelation of which is of interest to the agency. This concept has been used by states to reflect the complexities of large operations such as industrial parks and drinking water systems or perform other activities not limited to one site. These types of complex conditions are rare and difficult to capture in the data model.

The Facility Site Associations concept is not intended to capture associations that are based upon ownership, location, or industry, since they are already provided by the core data model. Facility Site Associations would apply only where the basic Facility Site data could not represent relationships between multiple discrete Facility Sites.

The states working with this concept have found very few examples to justify the additional structuring. All examples are limited to certain large, multi-tiered cleanup Facility Sites (e.g., a federal Facility Site). These complexities have been modeled via two approaches, both of which were found to be imperfect. The first uses a very broadly defined Facility Site as an umbrella and assumes that programs requiring better resolution could specify multiple Environmental Interests. The second approach establishes a hierarchy of up to three levels in which higher level Facility Sites represent associations of their corresponding sub-Facility Sites. Regardless of its level, each Facility Site has its own set of name, location, affiliation, etc., or inherits the information from its direct parent.

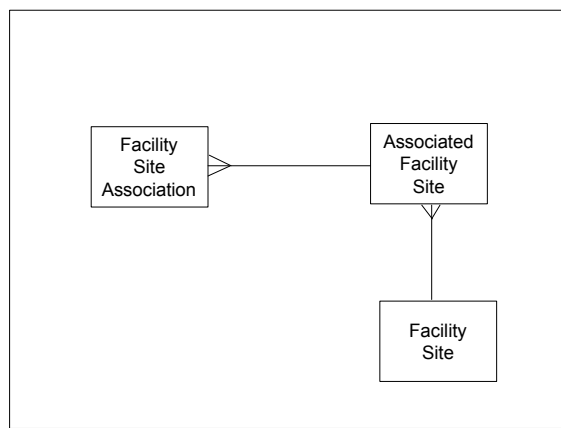
Neither approach has been used extensively. The first approach has not proven to cover all needs for Facility Site Associations. The second approach has not been applied broadly enough to justify the increase in system complexity.

The need to provide Facility Site Association capability may be more political than practical, particularly at the outset of a state's integration effort. Suspicious that an integrated agency data set cannot truly reconcile one's own specific definition of Facility Site with those of other users, programs are typically reluctant to adopt new protocols for Facility Site identification, especially for their most critical data. The suspicion may be temporary when it becomes clear that the data integration effort is not trying to change regulatory definitions used by the programs.

Nonetheless, providing some structure that accommodates more flexibility and variations for how programs link their data (i.e., via associations) may reduce such concerns.

The simplest approach to providing Site Association capability may be to create an entity representing a group of Facility Sites, with the ability to name and type such an association. In the figure below, a Facility Site can participate in more than one association. This modeling is based on limited experience and in itself is not known to be proven. States wanting to incorporate Facility Site Associations should first identify potential types of association and estimate the frequency with which they would be used. Unless this analysis yields types that are not already supported by components of the core data model (for example, ownership, industry, location) or are of high frequency use, the state should defer the inclusion of Site Associations. After the first release of its system, a better understanding of need for Site Association capability will emerge.

Facility Site Association Relationship



FACILITY SITE UNIT

A Facility Site Unit is a structure or specific point of interest at a Facility Site (e.g., a discharge point, a building, a landfill cell, a cleanup site). Programs collect information about Facility Site Units because they provide specific details about the location and nature of environment-affecting activities. Data on Facility Site Units is collected through such activities as permit development and regulated reporting.

Facility Site Unit has been excluded from the core data model because it is not a requisite during initial transition to an integrated system, and its integration and use are not yet well proven. Preliminary lessons learned about how this component might be structured are given for states that deem Facility Site Unit to be necessary. The data model diagram that represents these lessons is included for reference.

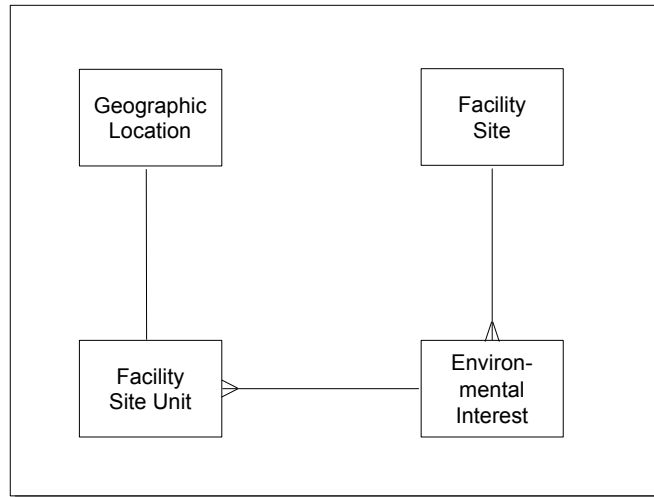
A Facility Site Unit should be described by a “type” attribute and should also include a name/description field. It is very desirable to record the location of a Facility Site Unit, especially by Geographic Coordinate. However, this level of detailed information (i.e., multiple coordinates within each Facility Site) typically is not collected. Significant state resources would be required to do so. If a program has such data, then it is worth performing a quality assessment of the data before migrating it into the agency system.

A Facility Site Unit should be recorded at the Environmental Interest level and not the Facility Site level. This recommendation is based on two factors. First, a program does not typically have a good way to verify that the Facility Site Unit with which it is concerned (e.g., a stack) is exactly

the same as that of concern to another program. For example, a Facility Site may have many stacks, and different agency programs may regulate a different subset of them. Second, Facility Site Units recorded at the Facility Site level would also have to be duplicated at the Environmental Interest level so that the program's environmental context is correctly reflected (i.e., is the tank at the Facility Site important because it is leaking, because it is licensed, or because it comes under the regulation of a wastewater treatment permit?). To provide an environmental context, the Facility Site Unit must be directly related to the Environmental Interest, which in turn is tied to the Facility Site.

When deciding Site Unit capability, developers should consider whether other detailed data that depend heavily upon Facility Site Unit would also be integrated. If the state is intending to include detailed permit information, then the Facility Site Unit capability would be highly critical to success.

Site Unit Relationship



AFFILIATION

Partial Owners of a Facility Site

Some states may wish to reflect the existence multiple, partial owners of a Facility Site and their percentages of ownership. The data model currently supports each Facility Site to having multiple Affiliations, potentially including the types Partial Owner and Complete Owner. An alternative approach would be to model a second entity called Ownership, with attributes of Type and Percentage of Ownership. The entity Ownership would have a mandatory relationship to Facility Site and a mandatory relationship to Organization.

Facility Site-Free Affiliations

The FITS II core data model assumes that Environmental Interest and Affiliation cannot occur separate from Facility Site. In other words, without an actual Facility Site, relationships to other entities cannot exist. In at least two states, however (both with centralized architectures, see next section), this assumption is not valid. In those systems, it is possible to have an Environmental Interest or Affiliation linked not to a Facility Site but to a person or corporation. In practice, the Facility Site-free Affiliation has been used to license or authorize a party to manage a multitude of identified cleanup sites. It remains unclear how this type of relationship might manifest itself in less centralized architectures.

Organizational Hierarchy

Some states require certain Facility Sites (e.g., mines) to provide information regarding their corporate hierarchy. If a paper record of this information is judged inadequate, then the core data model could be expanded to include a recursive many to many relationship. The entity Organization would require an attribute type to identify the nature of the relationship among each of the Organization's levels.

MISCELLANEOUS ALTERNATIVES

Tagging Facility Sites with a Federal Interest

Some states have expressed the desire and are currently exploring the added functionality of tagging Facility Sites with a federal interest indicator (especially reporting or jurisdictional issues). States are interested in discerning, for example, water discharges of interest to both state and federal programs from those of interest only at the state level. Through the use of an attribute linked to Environmental Interest, the user could mark if a particular program had a federal interest in that Facility Site or choose an interest type from a list of domains. At least three states are considering adding this capability. This functionality may become more important as states begin to transition to integrated systems which have business requirements to interchange data with federal systems.

Merging Geographic Location Addresses with Mailing Addresses

Some states model the Geographic Location Address attributes by merging them with the existing Mailing Address and using a specific Affiliation type for Location. Merging the two sets of address attributes avoids duplicating an address when the mailing address of a Facility Site is the same as its physical address and the data model may appear less complicated to understand.

However, the rules (and data quality requirements) for physical address data differ, in some cases, from those of mailing address and must be managed independent of the data model. For example, a physical address cannot contain a postal code and is sometimes little more than a description of the location, while a mailing address within the USA must contain a ZIP code. Mailing addresses also tend to change more frequently than physical addresses. If not carefully managed, the change could erroneously alter the physical address because they were recorded as identical.

The method by which location data might be processed to support system address matching capabilities (e.g., automatically converting an address into a geographic coordinate) may require a format inconsistent with that required by the US Postal Service.

Tracking Data Timeliness and Quality

Two factors that often drive an agency to integrate Facility Site data are the desire to streamline separate and potentially redundant internal information management efforts and enhance external use of the agency data. Both drivers heighten the need for the quality of data to be noted and understood by all users, inside and outside of the agency. The qualification that often provides the best insight as to the quality of data is its age.

When programs are mutually responsible for updating each other's Facility Site data, program users will often not understand the state of the data that may have been provided by another program. When a reporting form is received that reports on a Facility Site's status six months prior, and the user notices it represents different data from that which is already in the system, the user would need to know if the system data had been recently updated by another program or is older than that reflected by the report. External users may also draw incorrect conclusions if they

do not understand the timing of the different verification procedures required by various programs.

The nature of an integrated Facility Site system requires the agency to formalize a method for keeping track of data qualifications. Two possible methods are noted below:

- Periodically verify identification information for all Facility Sites and track when a Facility Sites responded. This process provides a consistent approach to data collection, consistency, and quality.
- Include fields in the system database to track who is updating data, when, and for what reason (e.g., permit renewal, inspection). This enhancement provides a ready flag for the benefit of other data users (and other data updaters).

Providing Staff Contacts for Obtaining Additional Information

The Affiliation information in the core data model is intended solely to record external Individuals and Organizations affiliated with a Facility Site. Some states have expressed a business need to include agency staff contacts.

Internal agency users may need to identify agency staff directly involved with a Facility Site (e.g., inspector, cleanup site manager). An entity called Staff, related to Facility Site or Environmental Interest, provides this information.

In addition, external users may need an agency contact for further details regarding a Facility Site, possibly a specialist in technical assistance or public outreach. Conceivably, one staff person per type of Environmental Interest or regional office would accommodate such needs. The agency approach will dictate the structure of the supporting data model. For example, if the agency appoints one employee to respond to all inquiries regarding 'Large Toxics Users,' the system would include a relationship between Environmental Interest and an Employee entity. The inclusion of staff contact may add to the complexity and overhead of the system.

TECHNICAL ARCHITECTURE OPTIONS

As the basic model described in FITS became widely accepted as a credible general approach, the most common follow-up question asked by both state and EPA practitioners was, “What architectural approaches and work flows have states used to actually implement and operate their systems?” These issues were addressed generally in the original document in the discussion of “Warehousing vs. Data Management” systems. Given this intense interest, the FITS II advisory group agreed that a more thorough survey and analysis of state approaches was needed to complement the update of the FITS model itself.

One of the virtues of FITS II is its balance between detail and overview. The model is neither a physical table structure, nor a formal logical table structure. No state would or could simply lift the model, code it, and drop it into their agency. Likewise the three types of architectures described in this chapter are simplified examples provided for purposes of illustration only. Because implementation of any architecture is subject to many more individual factors than the data model itself, no preferred architecture is identified. As described below, each state's architecture will be different, and all states will likely use some combination of the approaches outlined below. In addition, while the pace of development has generally eased and simplified the process of data modeling, the development of data management tools has totally revolutionized the implementation of data models. The illustrations below could be combined and implemented in a myriad of web-based and other application tools. Data warehousing and metadata management tools continue to proliferate, while at the same time, enterprise application development suites (e.g., those offered by Microsoft, Informix, and Oracle) continue to grow in sophistication. Beyond recognizing that most states are migrating to or planning to migrate to thin-client browser-based applications, this discussion does not cover the technical approaches used by states to implement these architectures.

HIGH LEVEL ARCHITECTURE

The FITS II model focuses on managing the relationship between two kinds of data: Agency and Program. The architectures used to implement the model in any given state system must therefore manage the relationship between these and two other aspects of the system: Agency Applications/Functionality and Program Applications/Functionality. Prior to state integration investments, most agencies built applications that were used exclusive to a specific program. Data needed by multiple programs was duplicated; often creating unmanaged repositories of inconsistent data. The FITS model represented a departure from this history, by designating some portion of each program's data as “agency” and building agency applications to manage it. For most states, this process has begun with the Facility Site data modeled in FITS and evolved to include other programmatic data like permitting, enforcement, and compliance. The key challenge faced by all systems is the management of critical relationships between program and agency data. Given this, it is useful to keep in mind four questions when reviewing proposed architectures:

1. What data is considered “Agency” and where should it be stored?
2. When a program creates or updates Agency data, (e.g., records a new permit) how does that information get into the system/database?
3. When Agency data changes, how are program systems/databases updated to remain current and consistent?
4. What ensures that #2 and #3 happen reliably? How is the Agency system/database kept current and consistent with that of the programs and vice-versa?

A review of state architectures suggests three general approaches to managing these relationships. The approaches are reviewed below:

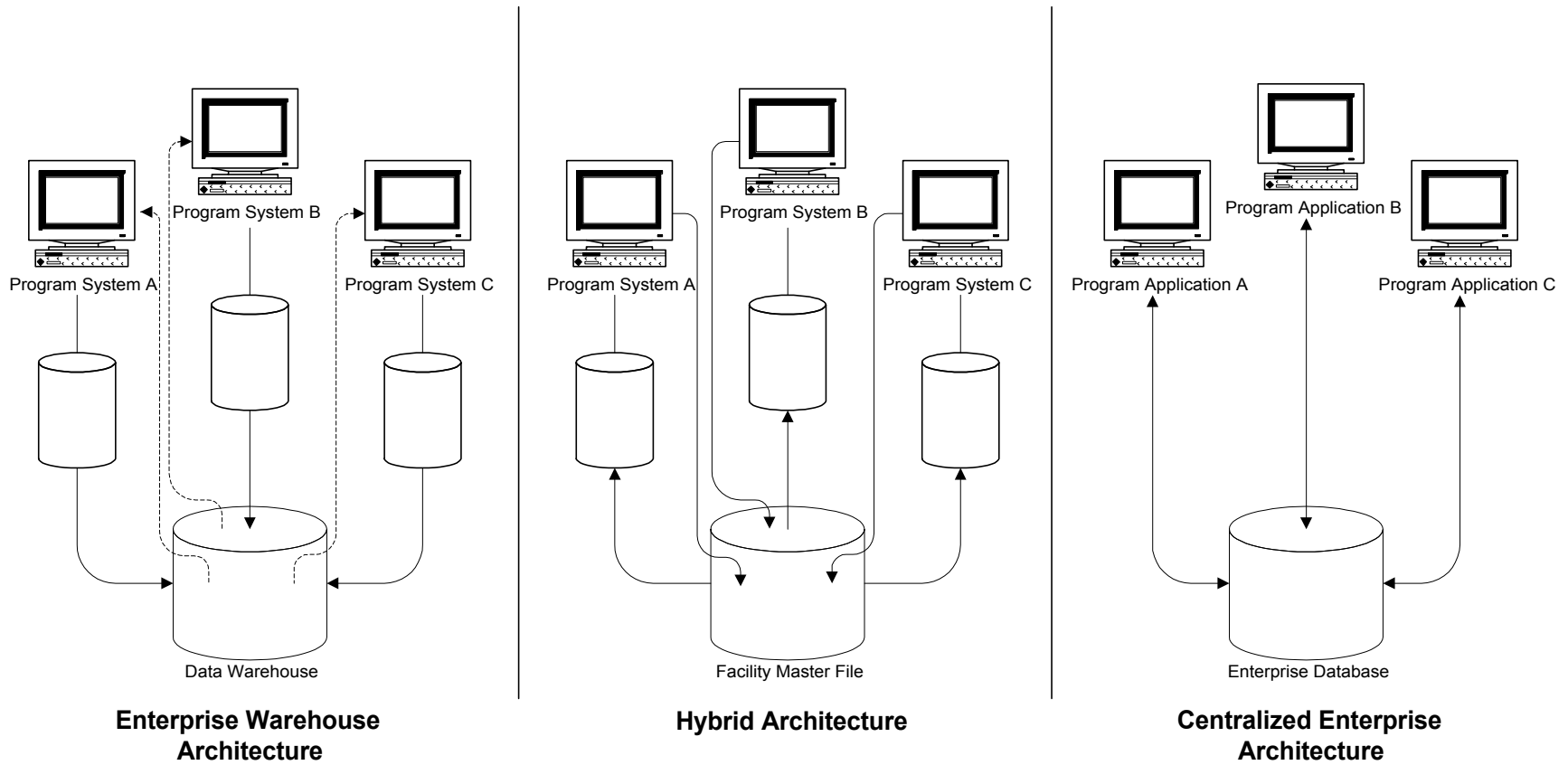
- Enterprise Warehouse:** A single enterprise, or agency-wide, warehouse stores a small core set of normalized core data and copies of some program data. Updates entered at the program level are pulled into the warehouse and reconciled. Updates entered at the agency level are not, however, automatically pushed to the program systems. Instead updates from the agency to programs occur through a passive notification and manual update process. Consistency of program systems and the warehouse is enforced by a combination of automated procedures and passive management processes.
- Hybrid System:** A single enterprise (agency-wide) database master file stores a small core set of normalized data and copies of some program data. The master file is updated from program applications (not databases), from an enterprise data entry application, or both. Updates to program systems are actively pushed from the master file. Consistency of master file data in program systems is enforced by an active, rigorous set of fully automated procedures and business rules, including the ability to override inconsistent changes made to program databases.
- Central Enterprise Database:** A single enterprise (agency-wide) database is the only authorized store for data, which is used by both agency and program applications. Updates and consistency checks and corrections occur within this single database across program and agency data.

These approaches are depicted in the diagram “Generalized Technical Architecture Approaches.” In the diagram, the program applications are depicted as computers, the central repository is depicted as the central large database, and the small cylinders represent an automated or manual process of data exchange. Enterprise applications are not shown. In the enterprise warehouse, the program databases feed (one way arrows) the warehouse. The dashed line from the warehouse to the program systems represents the management process used to update the program systems. In the hybrid architecture, the program systems are linked via bi-directional flow to the master file. This approach uses the master file to both read (e.g., find an existing Facility Site) and write information (i.e., record a permit application), then push data through a bi-directional link with the program database. This link also allows the master file to pull copies of program data. The central enterprise diagram shows that all data—program and agency—is stored and managed centrally.

The table below summarizes the key features of these approaches:

Architecture Type	Agency Data Stored In	Updates to Agency Data	Updates to Program Data from Agency Data
Warehouse	<input type="checkbox"/> Copy in warehouse; <input type="checkbox"/> Original in program system	<input type="checkbox"/> Program systems pulls only	<input type="checkbox"/> Management process – no automation
Hybrid	<input type="checkbox"/> In master file <input type="checkbox"/> Copy in program system	<input type="checkbox"/> Agency data entry application	<input type="checkbox"/> “Write-back” of agency data from master file to program system
Centralized	<input type="checkbox"/> Central database	<input type="checkbox"/> Agency data entry application	<input type="checkbox"/> Same data used for both, no update needed.

Generalized Technical Architecture Approaches



Each architectural approach has benefits and drawbacks. Two aspects are most important in differentiating these approaches:

- How does the architecture support the ongoing business process of maintaining a high quality, authoritative, timely agency repository?
- What implications does this architecture have for the autonomy of program and agency system managers?

These themes are at the heart of the very concept of agency information integration. Without significant management commitment and investment by program offices in the agency system, it cannot function. Program offices produce and use most of the data the agency system relies upon. The ongoing quality of agency data will suffer unless it is a priority to the program systems. Program staff can only work with Agency data if their applications allow them to do so. These differences are summarized in the table below:

Architecture Type	Unique Benefits	Unique Challenges or Costs	Co-management required between Program and Agency
Warehouse	<ul style="list-style-type: none"> ❑ Requires minimum co-management with program system administrators ❑ Can span many systems and platforms ❑ Can be very quick and inexpensive to build 	<ul style="list-style-type: none"> ❑ Consistency of program and agency systems totally dependent on passive management processes ❑ Program systems can operate with (and be unaware of) serious errors in agency data 	<ul style="list-style-type: none"> ❑ Minimal: once initial normalization and standards are negotiated, program systems need only allow “read” access for refreshes of warehouse
Hybrid	<ul style="list-style-type: none"> ❑ Allows existing systems to be operated with limited changes ❑ Can span many systems and platforms ❑ Provides high operational integrity of agency and program data 	<ul style="list-style-type: none"> ❑ Requires some modifications to program applications ❑ Interfaces are complex, especially with large numbers of systems and platforms 	<ul style="list-style-type: none"> ❑ Moderate: program systems must be modified to preclude direct changes to agency data yet allow “write-back” of agency data from master file
Centralized	<ul style="list-style-type: none"> ❑ Provides strongest possible enforcement of data consistency and quality procedures ❑ Provides base for development of broader programmatic integration ❑ Can be used to enforce and promote standardized business processes across programs 	<ul style="list-style-type: none"> ❑ Requires total commitment of agency and program staff to develop and use ❑ Central DB must be highly reliable and robust; maintenance and upgrades must be coordinated with all programs ❑ All program office applications must be modified to use central database 	<ul style="list-style-type: none"> ❑ Very high: program applications and databases must be developed centrally and tightly coordinated

GENERALIZED DATA AND WORKFLOWS

Technical architectures tend to manifest into common generalized data and workflows. Despite clear differences between architectures, all states conduct similar system and business processes, including the cleaning, reconciling, resolving, and updating of data. The architectures vary in where and how these essential processes are undertaken.

For each of the three architecture types, the generalized approach, data flows, and workflows are documented and diagrammed in the following pages. Representing not a particular state system but an assemblage of many systems with common characteristics, the diagrams are intended only to paint a high-level picture of flow, not to create a blueprint for system and business processes.

Each data and work flow diagram is broken into the following content areas:

1. Program Activities—Interaction of the program systems and the agency system
- 2a. Population Activities (Warehouse)—How new data is entered into a system
- 2b. Update and Resolution Activities (Hybrid and Centralized)—Process for entering and resolving new or updated data
- 3a. Resolution Activities (Warehouse)—Processes for reconciliation, resolution, and updates
- 3b. Quality Assurance and Quality Control Activities (Hybrid and Centralized)—Processes for ensuring quality of the data in the agency store.

Content areas for the Warehouse approach differ slightly from Hybrid and Centralized approaches. Each work element is discussed as a bulleted item under its respective content area.

Each diagram also depicts several clipboards, which that the data and workflow are governed by a designed process. The designed process may be manual, automated, or a combination. There are four types of clipboards:

- Reconciliation Protocols—Business rules for identifying duplicate facilities
- Resolution Protocols—Business rules for resolving duplicate facility information
- Data Standards—Enforced agency identified data standards
- Update Protocols—Business rules for updating new facilities.

Enterprise Warehouse Approach

Three content areas apply to the enterprise warehouse architecture: Program Activities, Population Activities, and Resolution Activities.

Program Activities

The box at the left side of the diagram represents the interaction of the program systems and the warehouse. In this architecture, the warehouse has no direct impact on the program systems. Data flows in one direction. It is pulled only from the program systems to the agency warehouse.

- Clean existing program data
If data used in the warehouse from a program system is noncompliant with the any agency data standards, the program would be responsible for updating the data and resolving any issues (e.g., the data were incorrectly captured in their data fields) or for researching the Facility Site to identify any missing information (e.g., no location information for a Facility Site is available in the existing system).

Population Activities

The shaded box near the center of the diagram describes a mostly automated process that occurs as data is pulled from the program systems into the data warehouse.

- Extract new Facility Sites
An automated process would extract the data required to populate the warehouse from each of the currently linked program information systems. Different rules would apply for each program system based on their data structures, quality, and rules. Programs may want to restrict some types of Facility Sites from being loaded.

- **Extract updates**
At a predetermined frequency (e.g., weekly, monthly) all Facility Sites currently populated in the warehouse would be re-extracted from the program systems and the warehouse data updated where necessary.
- **Verify/refine location address/coordinates**
A third party software tool would be used to validate and refine the location address data (e.g., converting to US postal standards, ensuring match for zip, city, county data) and to generate a latitude/longitude coordinate (if not already available) based the program's location data. For Facility Sites with inadequate data, a report would be produced identifying the data issues for resolution.
- **Add new Facility Site to warehouse**
Each Facility Site data record extracted from the program system would be entered into the warehouse. This procedure would happen only once per Facility Site, although a duplicate existing in another program system would also be entered into the warehouse, and the duplication resolved at a later time.
- **Add/update programmatic Facility Site data**
When either a new Facility Site is entered into a warehouse, or an existing Facility Site is to be updated in a warehouse, all of the programmatic Facility Site data (e.g., Name, Environmental Interest, Affiliation) would be added or replaced. If the Facility Site name or location has changed in an originating program system since the Facility Site was loaded, then a separate procedure (i.e., 'Process enterprise data update') would be initiated to ensure that necessary changes are reflected in the warehouse.

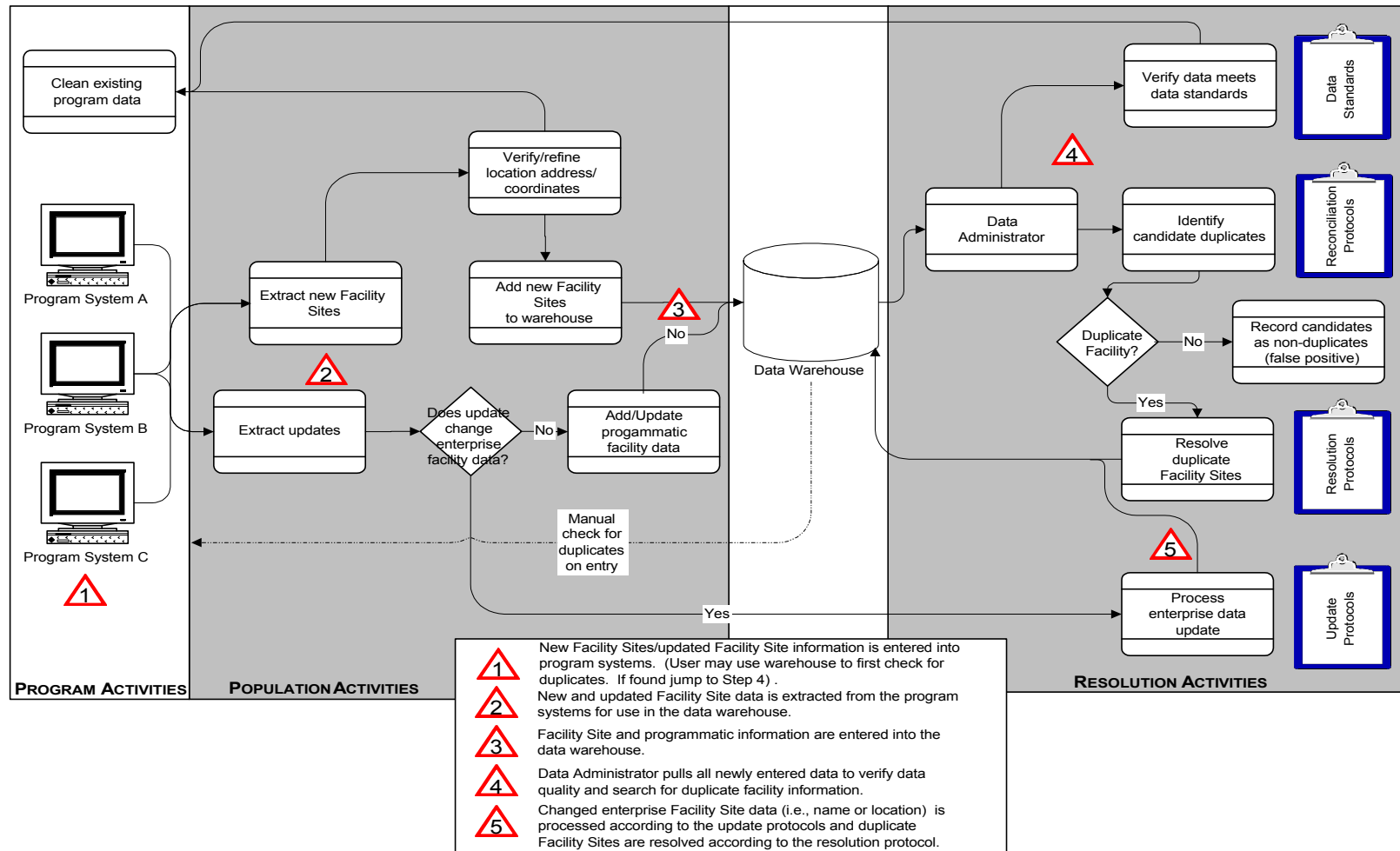
Resolution Activities

The shaded box at the right of the diagram describes the mostly manual process conducted independently of the data population activities to reconcile, resolve, and clean data in the warehouse.

- **Verify data meets data standards**
Quality assurance activities (possibly supported by automated reports) would scrutinize the data in the warehouse to ensure compliance with the minimum level of data quality required by the agency data standards.
- **Identify candidate duplicates**
Using automated support tools (e.g., a report of all Facility Sites with very similar names and in close vicinity to one another), all candidate duplicate Facility Sites would be identified and reviewed to determine if they are valid duplicates or just similar, but different Facility Sites.
- **Record candidates as non-duplicates**
Candidate Facility Sites determined not to be duplicates would be marked as previously screened, or clarifying information would be added to reduce or eliminate appearances in future duplicate reports.
- **Resolve duplicate Facility Sites**
Using the resolution protocol as a basis, a resolution team would determine how duplicate Facility Site name and location data should be resolved. The warehouse would be updated to include the results of this procedure.

- Process enterprise update
If a Facility Site name or location is changed in a program system and the Facility Site is already present in the warehouse, applicable information would be reassessed by the resolution team to determine necessary updates.

Enterprise Data Warehouse Approach



Hybrid Approach

The three content areas that apply in the hybrid approach (HA) are Program Activities, Resolution Activities, and QA/QC Activities.

Program Activities

The box at the top of the diagram describes the interaction of the programs systems and the Facility Site master file. All new or updated Facility Site data is entered at a single point directly in the master file and then immediately written back into the program systems. Data is not entered directly into the local program systems.

- Clean existing program data
If data used in the HA from a program system is noncompliant with the any agency data standards, the program would be responsible for updating the data and resolving any issues or researching the Facility Site to identify any missing information (e.g., no location information for a Facility Site is available in the existing system).
- Instant/Periodic write-back to program systems
Once the user has completed the business or system requirements in the resolution activities, the Facility Site master file would write back its data to the program system.

Update and Resolution Activities

The shaded box in the center of the diagram describes the process for entering and resolving new or updated Facility Site data.

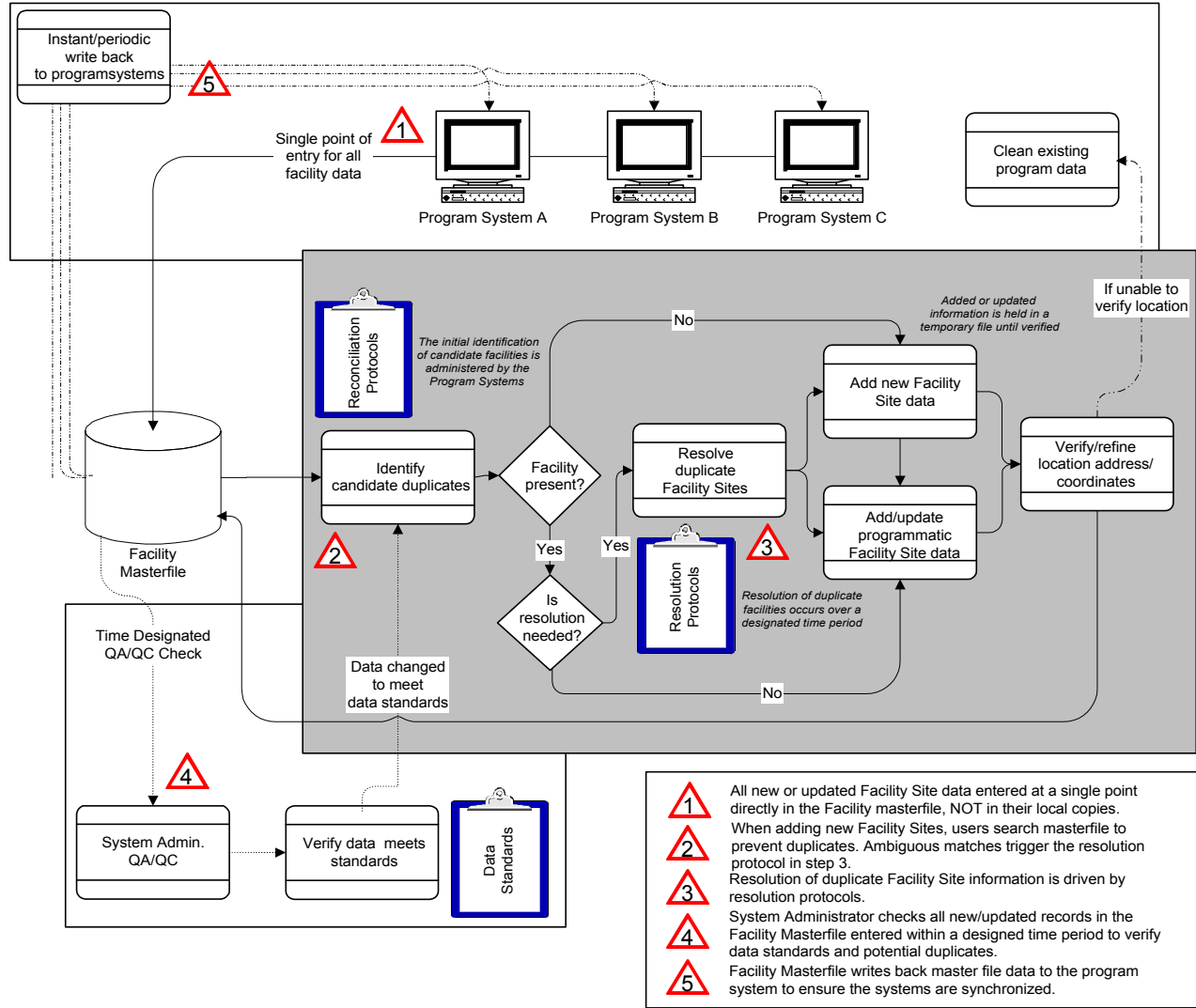
- Identify candidate duplicates
Before adding new Facility Sites, users would search the master file to identify duplicate Facility Sites. Using automated support tools (e.g., a report of all Facility Sites with very similar names and in close vicinity to one another) and manual searches, all candidate duplicate Facility Sites would be identified and reviewed to determine if they are valid duplicates or just similar, but different Facility Sites. Ambiguous matches would trigger the resolution protocols.
- Resolve duplicate Facility Sites
Using the resolution protocol as a basis, the interested parties would determine how the duplicate data should be resolved. The master file would be updated to include the results of this procedure.
- Add new Facility Site data
The user would enter new, resolved, or updated Facility Site data. This data would be entered through the single point and not directly from the program systems.
- Add/update programmatic Facility Site data
Although technically seamless to the user, when either a new Facility Site is entered or an existing Facility Site is to be updated in an HA, all of the programmatic Facility Site data (e.g., program Facility Site name, environmental interest) would be added, replaced, or updated. The default programmatic data would be the previously entered data, if applicable.
- Verify/refine location address/coordinates
A software tool would be used to validate and refine the location address data (e.g., converting to US postal standards, ensuring a match for zip, city, and county data) and to generate a latitude/longitude coordinate (if not already available) based the program's location data. For Facility Sites with inadequate data, a report would be produced identifying the data issues for resolution.

QA/QC Activities

The box at the bottom left of the diagram describes a manual and automated process of quality assurance and quality control.

- Time designated QA/QC check
The system administrator would periodically pull all new or updated Facility Site and programmatic information to ensure the data meets adopted standards and to identify possible duplicates. Candidate Facility Sites determined not to be duplicates would be marked as previously screened or clarifying information would be added to reduce or eliminate appearances in future duplicate reports.
- Verify data meets data standards
Quality assurance activities (generally supported by automated reports) would scrutinize the data in the master file to ensure compliance with the minimum level of data quality required by the agency data standards.

Hybrid Approach



Centralized Enterprise System Approach

As in the hybrid approach, the three content areas in the centralized enterprise system approach (CESA) are Program Activities, Resolution Activities, and QA/QC Activities.

Program Activities

The box at the left of the diagram describes the interaction of the program areas with the centralized system. Independent program systems do not exist. All data—facility and program—is kept in a centralized system and accessed from program portals. All new or updated Facility Site data is entered at a single point directly into the central database.

- Clean existing program data
If data used in the CESA from a program system were noncompliant with the any data standards, the program would be responsible for updating the data and resolving any issues or researching the Facility Site to identify any missing information (e.g., no location information for a Facility Site is available in the existing system).

Update and Resolution Activities

The shaded box in the center of the diagram describes the process for entering and resolving new or updated Facility Site data.

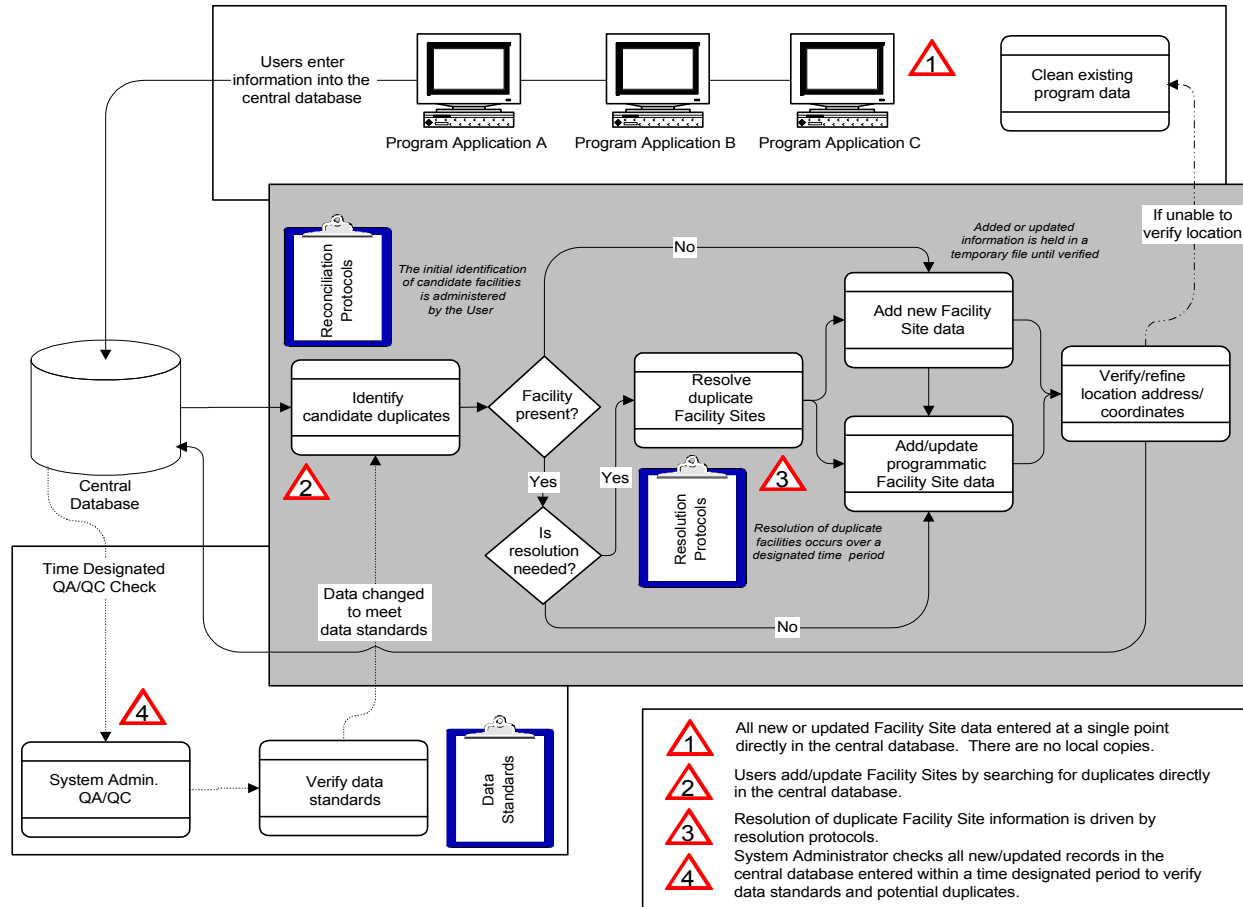
- Identify candidate duplicates
Before adding new Facility Sites, users would search the central database to identify duplicate Facility Sites. Using automated support tools (e.g., a report of all Facility Sites with very similar names and in close vicinity to one another) and manual searches, all candidate duplicate Facility Sites would be identified and reviewed to determine if they are valid duplicates or just similar, but different Facility Sites. Ambiguous matches would trigger the resolution protocols.
- Resolve duplicate Facility Sites
Using the resolution protocol as a basis, the interested parties would determine how the duplicate corporate data should be resolved. The central database would be updated to include the results of this procedure.
- Add new Facility Site data
The user would enter new, resolved, or updated data (e.g., name, location). This data would be entered through a single point.
- Add/update programmatic Facility Site data
When either a new Facility Site is entered or an existing Facility Site is to be updated in a CESA, all of the programmatic Facility Site data (e.g., Name, Environmental Interest) would be added, replaced, or updated. The default programmatic data would be the previously entered data, if applicable.
- Verify/refine location address/coordinates
A software tool would be used to validate and refine the location address data (e.g., converting to US postal standards, ensuring a match for zip, city, and county data) and to generate a latitude/longitude coordinate (if not already available) based the program's location data. For Facility Sites with inadequate data, a report would be produced identifying the data issues for resolution.

QA/QC Activities

The box in the lower left corner of the diagram describes a manual and automated process of quality assurance and quality control.

- Time designated QA/QC check
The system administrator would periodically pull all new or updated Facility Site and programmatic information to ensure the data meets adopted standards and to identify possible duplicates. Candidate Facility Sites determined not to be duplicates would be marked as previously screened or clarifying information would be added to reduce or eliminate appearances in future duplicate.
- Verify data meets data standards
Quality assurance activities (generally supported by automated reports) would scrutinize the data in the central database to ensure compliance with the minimum level of data quality required by the agency data standards.

Centralized Enterprise System Approach



TECHNICAL ARCHITECTURE INVESTMENT DECISIONS

Technical architecture decisions should fit the business need of the agency at a point in time. Architecture choices are not set in stone and the differences between architectures are not clear-cut. Although subtle differences exist, a number of states adopt characteristics from several architectures. In practice, the three technical architectures described in the previous sections are not simply three categories of approach; rather they represent three distinguishable points on a large spectrum of architectural options. States entering into facility identification integration should focus on what point on the spectrum best fits the business needs of the agency and subsequently how or if they wish to provide broader enterprise functionality in the future. Commonly, states expand from facility core data to the business areas of permitting, enforcement and compliance operation. States also expect their architectures to evolve, not locking a state into a single type of architecture. The exact trajectory a state chooses to evolve will depend on its own priorities, business needs, technology, experience and system lifecycles. Consequently, the initial decision on technical architecture is less important than the business analysis supporting the decision and future plans. Broadly, initial technical architecture decisions are based on the following factors:

- Program interaction and buy-in
- Funding and internal resources
- Consistent management and strategic vision
- Opportunity.

Program Interaction and Buy-in

A realistic assessment of program buy-in is critical to the success of any facility integration project. This section does not discuss strategies for achieving program buy-in. However, there is a strong relationship between the interaction of the agency and program systems and the potential resultant impact on program systems. This impact is often a key point of issue in assessing project buy-in.

Funding and Internal Resources

State must gauge the funding and available internal resources required to build and support an integrated framework. Many states choose not to initially develop centralized systems because the costs of implementation/purchase and the maintenance/upgrading of the system are prohibitive. Conversely, the enterprise warehouse is a lower risk, lower cost approach to achieving some increased agency functionality and benefits. Regardless of approach, resources must be allocated to support the agency system.

Consistent Management and Strategic Vision

Top-level management support and consistency with the overall agency vision and strategic planning is vital. Most states that have attempted enterprise integration projects without executive management level support have either failed or have not significantly realized agency benefits and functionality. Critical success factors involved with increased management support and consistency with strategic planning are discussed in detail in the following section, Management Considerations For Agency Information Integration.

Opportunity

The success of the original FITS document clearly confirms that if one state is working on a solution to a facility-related integration issue, many more are probably attempting to answer the same question. In that light, states have begun efforts to collaboratively build their capacity to jointly solve facility-related issues and develop needed applications. The ability of a single state

to join a collaborative network of states with similar goals and issues may reduce development costs and enable states to jointly institute more expensive, but more robust agency systems.

The differing lifecycle stages of the agency's major systems are also among the strongest influence on enterprise architecture choices. While a few agencies have elected to replace many legacy systems simultaneously with a single integrated system, most states have taken a phased approach where different systems have a variety of connectivity to the enterprise repository. Major agency applications undergoing re-design offer the opportunity to either create centralized or hybrid connections before the system is put into production. Conversely, some systems, either because of the platform, complexity, or throughput, may be impractical to connect through anything other than an enterprise warehouse approach. Most states indicate a desire to move towards increasing the functionality available to the agency through an enterprise repository, in many, but not all cases, this is envisioned as evolution of the agency architecture to a more centralized model. This is especially germane when applications are early enough in their design that the facility data tables themselves can be linked in the enterprise system. In some cases, this has even accelerated application development because the common tables provided a jump-start for the developers and for programs dealing with facilities and facility types already in the central system.

Since mature enterprise systems are still a relatively recent development for most states, there is simply not enough experience to suggest a generic preferred trajectory through or between architectures. Given how dependent such choices are on the individual agencies situation, there may not be one.

The clearest finding on examining the experience of states to date is a familiar one; success is more dependant on establishing a core set of enterprise facility data management *practices* that are credible and implemented by most of the agency. Managing system transitions is as much about agreements and management commitments to the underlying data standards and workflows as it is about the technical architecture. More importantly, agreements and practices lay the groundwork for any architecture. Any state that has worked through business and management issues, especially the initial reconciliation and resolution of their facilities, will be well-positioned to begin or transition to such a architecture.

None of the architecture examples discussed offers a *preferred* approach for any state. Technology is blurring the distinctions between approaches and increasing inter-operability. This trend is clear on agency *access* functionality, which is migrating to a browser-client intranet approach. This "enterprise portal" model bundles various access tools (i.e., GIS, browsing, visualization) into a common web platform. This approach is inherently distributed and multi-tiered. It is likely that any state could use some of each of these architectures as the operational back-end for any system.

MANAGEMENT CONSIDERATIONS FOR AGENCY INFORMATION INTEGRATION

This section provides an overview of the leadership, organizational, cultural, and business process improvements states should consider to support the design, implementation, and long-term maintenance of integrated information systems. The material presented here is a compilation and rewrite of Appendix C (Business Process Considerations) and Appendix D (Critical Success Factors) from the original FITS document. Since the release of FITS in 1997, a variety of users have requested additional guidance on the less technology-based challenges of enterprise or agency-wide integration projects. This was clearly stated by the participants of the September 1998 FITS Technology Transfer Conference in Seattle. The Knowledge Transfer Action Team of the State EPA Information Management Workgroup, sponsors of this updated document and the Seattle FITS conference, acknowledged the need for broad-based support of the challenges identified in this section. The Action Team is planning additional opportunities for knowledge transfer in this area.

Changes and enhancements to leadership, organizational cultural, and business processes are often ignored or overlooked in systems design and development. For most States, agency integration efforts represent a significant culture change. They can require changes to both business processes and the organization. The following material is based on lessons learned by a variety of states in their efforts to move forward on integrated information systems. This discussion is provided to generate constructive thinking and action, not as a comprehensive primer on management challenges. It is intended to apply to agencies in general, but may not address all issues relevant to a given agency. Agencies may not follow the suggestions exactly, but the discussion highlights known issues to be considered in agency planning. Moreover, some issues identified in this section cannot be fully addressed until a state is actively involved in the development of a system or its strategic planning.

CRITICAL SUCCESS FACTORS

Experience has offered several critical factors that will determine if a state's agency-wide environmental information integration effort can succeed. Some factors can be assessed during the planning stage while others come into play once the integration project has begun. Early awareness of these latter factors can help ensure that the project, its resources, and directions are properly aligned, and that the necessary support is available as they become critical to the project's success.

Strategic and Tactical Simplicity

Many successful state integration projects are driven by well-formulated business strategies that require integration to support them. These strategies, when clearly defined and articulated, are used to justify the short-term pain and inconvenience that information integration projects incur and provide a framework for creating incremental achievements throughout the process. An overall information integration strategy that is linked to the overall strategic direction of the agency can be useful in clarifying roles, defining options, and focusing resources. Any strategic framework should be as simple as possible—the more complex the strategy, the more likely it will flounder.

- A viable strategic plan can ensure that efforts are prioritized, well founded, coordinated, and manageable. Some strategic plans are no more than a list of unconnected tasks, each of which appears to satisfy one of a multitude of vague goals. Such plans are neither cohesive nor

communicable, nor do they reflect the sense of priority that is critical to sustaining funding and staff commitment. A viable strategic plan should be 50% content and 50% motivational. Although it may address a period of several years (e.g., 3-5 years), the plan typically should provide greatest focus for a shorter timeframe (e.g., the first 1-2 years). After this initial time period, the longer-term elements can be fully revisited with the perspective offered by the changes of the initial years.

- Programs focused on a specific function or environmental medium traditionally are measured against program-specific achievements, and so program management and staff often see limited value in participating in, and compromising for, agency-wide projects. Clear links between integration project objectives, comprehensive agency-wide goals, and direction are important to articulate in any strategic planning product.
- While overall information integration projects may be large, the individual systems re-engineering steps should be small and contained. No single project should exceed 6-9 months in duration, so that interest, attention, and momentum can be maintained. Each project should deliver some increment of value by itself, even as it often takes many small projects together to realize significant global benefits to an agency. The absence of short-term, incremental benefits can jeopardize staff and budget commitment to the larger integration project.
- Tactical planning should account for the unexpected. For example, agency budgets and priorities can change significantly over time (e.g., when an administration changes) and, when they do occur, can cause delays in projects. The possibility of such impacts should be accommodated in planning and should not leave the agency exposed to incomplete information systems or to having set, and failed to meet, unrealistic expectations. In other words, the underlying assumptions and planning factors should be documented and communicated so that any changes can be understood and explained.
- The first project conducted to support an overall integration effort is the hardest. As such, it should be very small in scope, because most of the effort involved will be in learning a new way of thinking (i.e., cross-program) and in decision-making rather than in analysis. Nonetheless, the project should be large enough to still provide an increment of value to agency staff. As such, an initial project plan should be crafted that clearly identifies goals and objectives, timelines, resources, and deliverables. It can offer up the management structure and decision-making mechanisms/processes for the expected and unexpected.
- Integration projects across many programs should not try to integrate all of the data used by all of those programs. Some of the data is, and should be, shareable. Much of the data, however, is program-specific, and the individual program should be solely responsible for it. When in doubt, the agency may do well to assume that the information is program-specific. It can always be integrated and shared later, if necessary.

Leadership and Commitment

Agency-wide integration projects are more successful when strategically and tactically managed. These projects are usually one of the first enterprise projects taken on by an agency and are generally more complex than any program-specific projects. They require multiple programs to coordinate efforts and resolve issues that may conflict with more narrowly defined program practices. As such, executive and middle management commitment is critical to ensure that program-specific priorities incorporate new agency-wide objectives.

Executive and middle management involvement, visibility, and accountability are critical to project success. Programs that have approached enterprise integration as a purely information technology project have generally suffered poor results.

- Successful projects have generally had at least one executive or management team member that is personally and professionally committed to the project. This project champion,

preferably not the functional project manager, is responsible for facilitating the resolution of the variety of agency-wide issues that always emerge. In some states, this champion has been a Chief Information Officer accountable to agency leadership. In other states, a new management position (non-technical) with business knowledge has been created solely for information integration. In a few successful states, the champions have been at the commissioner or deputy commissioner level.

- Central budget management has been a factor in the success of many state integration efforts. The good will of program managers cannot be assumed. Taxing media or functional programs—that is, drawing project funding from individual program budgets/grants—is one way states have garnered program participation.
- Program commitment must continue beyond the end of any project. Information systems often fail due to a lack of continuing support from management, users, or technicians. The project budgeting should take into account operation and maintenance needs. Systems designed for multiple programs require more maintenance and support (e.g., help desk, training, coordination, potential program business changes) because of the greater number of users and the greater diversity of needs.
- Staff positions dedicated to these projects must be provided. Staff positions create institutional memory for projects whose results may take years to fully realize. A project manager is no longer a bare minimum requirement. A nontechnical, business-oriented data administrator position should also be staffed.

Project Management

Agency-wide integration projects are susceptible to risk. Typically, no individual in an agency understands the details of every program. This means that no one can easily predict the range of issues that the project will need to resolve. As a result, a sound project management structure needs to be established from the outset. This management structure needs to be able to address and resolve the issues that will arise during the process.

- A full-time, experienced project manager must be assigned up front. The project manager should have clearly defined authority and direct access to senior agency decision-makers.
- The project manager should understand the issues related to system development, business process re-engineering, data cleaning and migration, and system integration so that these efforts are all coordinated and realistically planned.
- The schedule of the integration plan is critical. It should be rigorous but realistic. Delivering products on schedule is critical, particularly to maintain credibility for projects that serve longer-term goals but consume short-term resources and that may not appear to directly satisfy short-term needs.
- A detailed project plan is vital, as is a specific project scope. The project scope should clearly define not only what is within the scope but in particular what is out of scope. Agencies should be wary of the inclination to saddle other, not strictly relevant, needs onto the information project.

Formal issues management is critical. Issues need to be identified and defined, and their resolution documented and communicated. The lack of rigorous issues management and communication will cause issues to be endlessly revisited and will create confusion and frustration about what is happening in the project.

- A formal project organization is critical. As noted above, successful integration projects have been chartered as an agency project, not as an information management projects. Successful structures have exhibited the following characteristics:

- Programmatic oversight representation from across the agency including technical and administrative support organizations
- Visible and communicated accountabilities to agency leadership
- Defined relationship with information technology support organization
- A decision process, with decision timeframes, defined up front. When programs are unable to make a decision, then someone, or some group of people, must be in place to make the decision for them.

Successful integration project organizations generally contain members who focus on the business policies and procedures to ensure that they will change to support a more integrated information philosophy.

The program staff with the least time available to participate in information projects are typically the very ones who can best define the information needs of the program and whose participation is most valued. Field staff have different philosophies and information needs than centralized staff. They must be integrally involved in any information decision-making.

- Most decisions (e.g. the definition of “facility”) are neither right nor wrong but are simply a choice of direction. A stalemate over these choices, which frequently occurs, is always counterproductive. A choice in approach that turns out not to work ideally can be revised. The key is to keep moving and not become paralyzed by indecision. If left unchecked indecision can derail a project.

Communication

Agency-wide integration projects demand ongoing involvement and support by all levels of staff and all agency programs. Successfully engaging staff and programs requires high quality, frequent communication. The absence of such communication is rated as the number one reason why enterprise-wide change projects fail.

A vision of long-term goals, developed through discussions with staff across the agency, has shown to be valuable. The vision should be as meaningful and tangible as possible. This goal setting is achieved by focusing on the future capabilities that the project will provide staff (and other information users, such as key stakeholders). Furthermore, any vision should be as concise as possible. If an integration vision cannot be explained clearly to anyone in less than 2 minutes, it is of little use.

- All agency staff must be updated on project status and decisions on a regular basis. Face-to-face communication is often the only way to ensure that staff members hear the message.

Technology

Although the benefits of integrated information systems are data and business procedures, and technology, a few general principles are critical for ensuring that technology supports information integration.

All staff must have convenient electronic access to agency information systems. The lack of easy access to systems and the information therein is a key reason for the breakdown of systems. The technology convergence on the Internet has allowed many programs to develop easy access browser-based front ends to more complex database systems. Most staff and stakeholders are becoming familiar with the ease and functionality of these popular applications.

Successful agencies generally have some type of central information systems coordination group that is accountable and empowered to establish and maintain technology standards. Technology decisions are now appropriately considered strategic and should take into account the long-term

view and establish discipline. Without agency-wide accountability, guiding standards and vision, technicians and developers may be persuaded to install the latest and greatest, which can significantly disrupt any agency-wide integration efforts.

- Historically, state environmental agencies have not been the type of organizations that take on cutting-edge technology projects. One reason has been the inability to retain highly skilled information technology specialists. However, the evolving convergence of technology options, the reduced development complexity, and the increased vendor stability allows agencies to consider significant technology leaps. Notwithstanding the ability to meet functionality, linkage, and access goals, as an agency intends to take significant technology leaps (e.g., moving from mainframe systems to client-server technology or Internet systems), the project should minimize risks by using pilot projects. Pilot projects allow a project to test new technology and approaches, implement changes incrementally, and secure experts to support these efforts.

Business Processes and Decision Making

The most significant change (and therefore challenge) required for information integration occurs at the business level, not the technology level. Most information systems fail because they do not meet the needs of business (i.e., program) staff. The most successful agency-wide integration efforts take place in organizations that can integrate their relationship with information and information management. These organizations succeed because they are willing to make the tough decisions that enable them to succeed. The remainder of this section provides an overview of business process improvements--the changes to policies and procedures necessary to support the design, implementation, and long-term maintenance of integrated information systems. This discussion complements the description of the FITS II core data model.

OVERVIEW OF BUSINESS PROCESS IMPROVEMENTS TO SUPPORT INFORMATION SYSTEMS

Business processes refer to agency procedures and practices that relate to data collection, quality assurance, entry, and use, maintenance of system content, and alignment of agency staff.

Staff involved in agency-wide integration efforts frequently observe that the agency's ability to manage the business processes for implementing and maintaining integrated information are more important to the project success than are the technical aspects of the system. It is one thing to design a system well technically. It is even more challenging, from an organizational perspective, to maintain a viable system that draws on data from across an agency and delivers current, high quality, consistent information users can count on. Individual programs can be expected to take a direct and involved role in managing the business processes needed to support their own systems. For agency-wide information integration, however, agencies need to recognize that there is likely no one clearly authorized to manage and enforce agency policy and procedures needed to support an integrated information system. As noted above, agencies need to anticipate and plan for policy and procedural issues and create the corresponding organizational capacity to support the successful design, implementation, and maintenance of the integrated system.

As agencies consider these issues, it is important to remember that the process is iterative. Decisions made about the intended system functionality (e.g., to provide a unique Facility Site profile) will influence decisions about design (e.g., a shared integrated database), which will in turn influence business practices (e.g., ensuring that programs follow agreed-upon naming conventions for Facility Sites). Once the system is in place, the actual conversion and use of legacy program-specific data (data generally housed in older programmatic systems) may require one or more of the original policies or decisions to change. Additionally, business needs may

change over time and require changes in the system. This process is normal and essential to keep the system current. It illustrates why the agency's decision-making process needs to be managed. It also speaks to the organizational requirements needed to support decision-making and the ongoing use of the integrated system. At a minimum, an agency will need a cross-program team to help with system development. Agencies further will need to coordinate ongoing system population and maintenance and provide the appropriate organizational arrangements to do so.

The business process improvement considerations described below are presented chronologically, generally in the order they would occur when undertaking an information integration effort. These points therefore may be useful as an additional reference for state planning and development work. When planning for systems integration, an agency needs to take into consideration agency policies, agency procedures, and program procedures:

- **Agency Policies:** the broad policies that the agency establishes to govern and define the direction, objectives, authorization, and management of the integration effort. At a minimum, these policies should address the data scope to be integrated for the agency, the standards for data quality, and a mandate for participation in the integration effort. Implementing these policies leads to decisions about the more detailed agency and program procedures, as follows.
- **Agency Procedures:** the procedures to be implemented agency-wide in order to successfully implement and maintain the integrated system. They should address issues such as how data should be converted into the integrated standard, how new Facility Sites are to be recognized and entered into the system to avoid duplicates, and how frequently data quality and content are to be verified. The agency procedures to be established will influence the procedures used by the individual programs, as follows.
- **Program Procedures:** the changes or additions to existing practices necessary to ensure the programs can adequately support the integrated system. Agency procedures must be translated into basic workflow descriptions for individual programs, so that the programs can understand how to modify their individual practices to support the agency requirements.

Agency Policies

From a procedural standpoint, agency policies are needed in two key areas: scope and data quality.

Scope

The agency must determine the breadth of the data to be integrated, maintained on behalf of, and shared by the agency as a whole. Individual programs solely manage whatever remains. As practical examples, the agency will be considering such questions as the following:

- For Geographic Location, will the agency reconcile the programs' data to establish a single physical address for the Facility Site, or will the programs maintain their own physical address information, instead of or in addition to an official agency physical address?
- For Industry Classification, will the agency establish a primary industry code for each Facility Site, or will the agency allow multiple industry codes to characterize industrial activity at the Facility Site? Will multiple industry codes be supported for each program?
- For Affiliation, will the agency reconcile any of the Affiliations, such as owner information, and maintain it once on behalf of the entire agency? Will the agency allow the programs also to maintain their own owner data, or data about other Facility Site contacts specific to the program?
- For Organization, will the agency try to uniquely identify each Organization?

In determining the scope of the integration effort, the agency should take into consideration the following factors:

- Burden and resources—some data managed by individual programs about Facility Sites of environmental interest is redundant when considered across programs. Economies of scale and a reduction of burden (externally on data reporters, and internally on data managers) may be achieved by consolidating data management responsibilities at the agency level.
- Usability—the business needs should define what the data content needs to be in order to be usable as intended. Getting the necessary content may require new procedures to collect the information.
- Access—users, especially program staff, need to be able to access information readily, and specific access requirements will have implications about who manages the data and where and how it is managed. The answer to these questions will drive organizational decisions about roles and responsibilities to support and manage information.

Data Quality

The agency must define what quality means for its integrated data. The quality objectives will guide how the system is designed; whether and how agency and program procedures are changed; how the data is transitioned to the system; and how the system is maintained. Quality objectives can only be achieved by establishing adequate procedures across the agency, within the individual programs, or both.

Data quality can be considered to include coverage, timeliness, accuracy, and understandability. Each element may result in different implications for agency and program procedures.

Coverage

Does the agency have the desired data about all of its Facility Sites? For example, while many agencies would like to understand a Facility Site's industry, very few media programs currently collect SIC data for all of their Facility Sites. If the agency decided that the integrated Facility Site data must include SIC data, then data collection procedures would need to be created or modified. Additionally, if data points have no value, business practices may be changed to eliminate their collection and maintenance.

Timeliness

The agency needs to decide how current the data needs to be in order to be usable and worth the integration and maintenance costs. Some data may need to be more current than other data. For example, the agency might decide that SIC information only needs to have been verified within the last 2 years (the agency might consider it unlikely, for example, that a Facility Site's industry will change very much, very often). Owner information, on the other hand, might need to be verified at least annually because the agency needs to feel confident that it knows the responsible parties at a Facility Site. The agency's policies about data timeliness will drive decisions about how information is collected and verified, either by the agency overall or, by devolution, by the individual programs to meet the broader agency requirements.

Accuracy

The agency needs to consider how accurate the data needs to be. One obvious reason for integrating Facility Site identification information is to establish a single core set of accurate identification data. Here, too, the agency's answer about accuracy is determined by need. For example, the agency might decide that it is very important to have reconciled owner and

geographic location information in order to uniquely identify a Facility Site, but less important to have reconciled Facility Site contact information. If that is the policy, the programs could choose to retain their own specific Facility Site contacts. The agency's policy on data accuracy therefore will drive policies and actions for data reconciliation.

Understandability

Integration involves drawing information from various sources and program perspectives, reconciling it, and delivering it back to the programs and others. Integrated information therefore is supposed to support a fairly diverse audience of users—e.g., staff and managers across the agency, as well as agency stakeholders. Because of this broad audience, the agency needs to invest in making the integrated information as understandable as possible. This means:

- Establishing standard ways of reconciling and presenting the information with a common look and feel, for example standard naming conventions, standard abbreviations (does the abbreviation SW in the system refer to solid waste or surface water), and standard coding systems
- Providing metadata (i.e., data about data) for the users' benefit to understand how the data were developed and their timeliness and quality.

Implications of Scope and Data Quality Policies

Scope and data quality decisions are highly interrelated. The agency may find that all data of interest to the agency is either not collected or is not collected regularly by some programs. Therefore, data quality (in terms of coverage and accuracy) may not meet the agency's expectations. The agency's policies must influence how these situations are handled. For example, the agency will need to consider criteria such as the following:

- Data with less than X percent (e.g., 30 percent) coverage will be excluded from the system
- All existing data will be included as is, regardless of the coverage
- Programs will begin collecting the data to ensure that there is 100 percent coverage for all Facility Sites of environmental interest
- All Facility Sites must be reconciled.

Agency Procedures

The governing policies that the agency establishes regarding scope and data quality will dictate changes in procedures, either at the agency level or the individual program level. This section describes possible implications for agency procedures.

Four main types of procedures are likely to be affected by agency data integration policies:

- New data collection and entry
- Data conversion
- Ongoing quality assurance
- Training.

New Data Collection and Entry Procedures

How will new Facility Sites be entered into the system and how will existing Facility Site information be updated? The agency's overriding concern is to prevent the entry of duplicate

Facility Sites. A second need is to maintain the quality of the data through the life of the Facility Site. Therefore, any updates must be controlled according to procedure.

For example, to prevent duplicates, the agency could establish a series of rules to follow when a new Facility Site is to be entered. For example:

1. The newly entered Facility Site must contain at least the minimum data requirements (e.g., name, geographic location).
2. The system should first be searched for the Facility Site name and any variations thereon. These would be scrutinized for possible duplication.
3. The system should then be searched for Facility Sites in the vicinity of the would-be Facility Site's physical location and any possibilities scrutinized.
4. If a possible duplicate emerges, programs with interest in the possibly matching Facility Site(s) should be consulted to resolve the question.
5. If consultation among the programs does not resolve the matter, the Facility Site should be called directly and consulted, and so on.

Similar procedures could be established for updating Facility Site information—i.e., defining what information should be updated. An important element of this process is determining who is responsible for the update and who approves the updated information. The programs are likely to be very sensitive about this, since there may be program-specific reasons for updating, or not updating the data, and the programs will want to know precisely when and how the integrated Facility Site records are being changed. States have handled this differently. For example, one state developed a hierarchy of its programs, based on which programs had the most rigorous statutory and regulatory authority about Facility Sites (e.g., the ability to withdraw a permit). In that state, when one program wants to update core information, it has to get the approval of whatever program has an interest in the Facility Site and is highest up the hierarchy of program authorities. In other states, if one program wants to update Facility Site information, it sends e-mail to the other programs with an interest in the Facility Site to discuss the update and agree on it. A database administrator intervenes if the programs cannot resolve the matter.

The agency may wish to establish a formal process and schedule for data verification. The frequency would be based on how current the data needs to be. A Facility Site verification report is one mechanism that could be used. Each Facility Site would be provided a report of the identification information currently held in the agency's system. The Facility Sites would verify the data and submit any corrections to the agency. Such a verification report could standardize how Facility Site information is collected and could pre-empt individual, uncoordinated, and potentially inconsistent program inquiries.

To ensure data quality, new Facility Site records and updates need to be guided by common understanding. For example, when recording the start date of an affiliation, is the program supposed to find out when the individual in question began working at the Facility Site (which would be difficult for an agency to verify), or is the program supposed to enter the date when the staff became aware of the affiliation? Such procedures ensure the quality and consistency of the data, particularly so users can understand what the data represent.

Data Conversion Procedures

How are legacy program data to be converted to the new agency standard for integrated data? The agency needs to define at a high level the steps that each program is to use to convert its data, although the programs themselves will tailor this direction for their specific data flows,

as described below in the “Program Procedures” subsection. The agency as well may provide specific tools for how programs convert the data, such as establishing a data migration engine. Finally, the agency needs to consider how to ensure and certify the quality of the converted data.

For example, the physical address data from each of the programs might need to be converted into the new standard format. The agency would need to guide the activity to ensure that the outcome—the converted data—meets the agency’s policies for scope and data quality. The agency might, for example, determine that physical addresses must be associated with latitude/longitude coordinates for the Facility Site. The agency would need to establish the preferred procedures for converting physical address data to include latitude/longitude. For example, use address-matching software; if that does not work, use a map; if that does not work, call the Facility Site; if that does not work, use GPS (Global Positioning System) tools, etc.

Once the data have been converted, a series of formal checks can be performed to verify data content and quality. The agency’s conversion procedures are designed to uphold the agency’s policies for the integrated data.

Ongoing Quality Assurance Procedures

The agency should develop ongoing procedures to measure data quality and content. This quality assurance is important to ensure that established procedures are being followed over time. It is particularly important to determine whether any of the policies and procedures are inadequate. Ongoing quality assurance procedures could take many forms:

- Conducting periodic (e.g., monthly) searches of the integrated information system to identify possible Facility Site duplicates
- Reviewing all records created or modified within a specified period of time to ensure that the attributes have been populated correctly
- Sending a periodic verification report asking the Facility Site asking to verify and correct the identification information in the agency’s system.

Ideally, an agency would create a cross-program team responsible for overseeing the ongoing maintenance and operation of the system. The team would be responsible for identifying issues and problems about how the data are being maintained, so that emerging problems can be fixed before they undermine the viability of the entire system. The team could also present changing business requirements for evaluation and resolution.

Training

A critical need to support the maintenance and use of the system is training. On the front-end, for data entry and support to the system, selected staff should be trained and certified in data entry and quality assurance procedures. Such training would maximize the likelihood that data are being entered consistently, according to established data quality rules. Quality assurance staff should be trained to monitor data quality on an ongoing basis. This training might include knowing how to verify the content and quality of recently entered Facility Site records, knowing how to search for and divine duplicate sites, and knowing how to analyze different elements of the Facility Site data to assess overall data consistency and accuracy and the adequacy of the business rules.

On the back-end, system users need to be trained on how to use the system, analyze the data within it, and interpret the attributes and values in the system. Particularly because the

integrated system is to address a wide audience (as opposed to, for example, a program-specific audience), training will be critical to ensure successful use of the system. The success with which staff are able to use the system will directly determine the extent to which resources and focus continue to be devoted to maintaining its data content.

Program Procedures

For the ongoing use and maintenance of the data, programs need to understand how traditional procedures need to be refined to work with the new agency policies and procedures. For example, the program may need to revise its permit application forms to collect the Facility Site identification information necessary to fully populate the Facility Site records in the integrated system. Some programs might create new Facility Site records when an owner changes, whereas the agency policy may be simply to record historic ownership for a Facility Site. The program would need to change its procedures to prevent creating a new Facility Site just because the owner changed.

APPENDIX A: DETAILS OF THE FITS CORE DATA MODEL

The following text presents the details of the FITS II core data model. This text is intended particularly for the benefit of the more technically inclined reader. For each component of the data model, as presented in the “Overview of the FITS II Core Data Model” section of the document, the text presents the following information:

- Entity definitions;
- Attribute definitions, usage, and examples; and
- Attribute format tables.

Again, the data model components have been organized as follows:

- Facility Site, Environmental Interest, and Alternative Identification components have been addressed jointly because of their close interrelationship
- Geographic Location
- Industrial Classification
- Affiliation

FACILITY SITE, ENVIRONMENTAL INTEREST, AND ALTERNATIVE IDENTIFICATION

Entity: *Facility Site*

Definition: A place where regulatory activities of interest to the agency occur, or have occurred in the past.

Usage: This definition is intentionally vague. Although the vast majority of Facility Sites are easily identified (e.g., a factory), some are harder to discern. This is partly due to the complex structures of some Facility Sites (e.g., non-contiguous land, multiple buildings) and to the varying and sometimes conflicting definitions specified in environmental rules. Experience has shown that, even for the more complex cases, Facility Site identification can be achieved, through a case-by-case approach, if necessary.

Relationships between Entities:

- A Facility Site is always monitored due to one or more Environmental Interests
- A Facility Site always is found at one Geographic Location
- A Facility Site sometimes performs activities associated with one or more Industry Classifications
- A Facility Site sometimes involves one or more Affiliations

Attribute: State Facility Identifier

Definition: A number randomly assigned to the Facility Site, which serves as the unique (‘key’) identifier of the Facility Site.

Usage: This number should be assigned as soon as a new Facility Site is identified by the state. Care should be taken to ensure that the Facility Site is actually a new Facility Site, so that duplicates of existing Facility Sites are not mistakenly created. If the Facility Site is correctly identified as being a new Facility Site, then the state facility identifier should stay with the Facility Site and should be used in all interactions with the Facility Site. The number assigned should be totally unintelligent (i.e., a random number). This ensures that it does not

provide any meaning that could become invalid if the source of that meaning should ever change.¹²

Examples: 434221523
00543245

Attribute: Facility Site Federal Identifier¹³

Definition: The non-intelligent, unique identification number (UIN) that will be assigned to every Facility Site by EPA once the program for doing so is completed

Usage: EPA intends to assign every Facility Site a UIN to facilitate data sharing with states. While not mandatory, the core data model recommends that states accommodate the federal number in order to minimize confusion when sharing and updating data with EPA.

Examples: 763243434213
865327543245

Attribute: Facility Site Name

Definition: The name assigned to the Facility Site by the agency.

Usage: The state will need to establish rules about what is an acceptable name for a Facility Site (e.g., the most visible name at the Facility Site, the name that the owner of the Facility Site uses for that Facility Site, etc.) and what the conventions for the format of Facility Site names should be (e.g., how abbreviations should be used).

Examples: Fairfield Smelters
Texaco Station #1553

Entity: *Environmental Interest*

Definition: An Environmental Interest is a need or activity that causes a program to be interested in a specific facility site; it is usually related to a regulation or regulatory program that applies to a Facility Site (e.g., Superfund site, Title V permit).

Usage: Environmental Interest provides the agency with the ability to track details (e.g., contacts, permits, status) about a Facility Site that are only relevant due to the interaction provoked by the Environmental Interest.

Relationships between Entities:

- An Environmental Interest always represents an interest at one Facility Site
- An Environmental Interest sometimes requires a program to interact with one or more Affiliations
- An Environmental Interest sometimes has one or more Alternative Identifications
- An Environmental Interest sometimes tracks one or more Industry Classifications for a specific Facility Site

Attribute: Environmental Information System Identification Number

Definition: The primary identification number used to identify a unique single environmental interest in a specific facility site tracked by that Environmental Information System. Additional identification numbers for a given Environmental Interest may be stored in the Alternative Identification entity.

¹² For additional information on the process for identifying new Facility Sites and ensuring that no duplicates are created, please refer to the Technical Architecture Options sections of this document.

¹³ The Facility Site Federal Identifier attribute in FITS II is equivalent to the Facility Registry Identifier attribute in the State/EPA approved Facility Standard.

Usage:	Usage: In some cases this number may simply be the permit, license or program-specific facility number used as primary identifier for that interest or facility in the source system. In other cases this number may be a separate, system maintained ID which allows multiple individual permits or other IDs to be managed as a single environmental interest. Some users may find this number itself meaningful, but in most cases it will be used only as a link to the program information (or information system).
Examples:	500674523 39027943205
Attribute:	Environmental Interest Facility Site Alternative Name
Definition:	A name other than the Facility Site Name that the agency would like to have affiliated with the Facility Site, such as a program-specific name for a Facility Site.
Usage:	It may be desirable to record a Facility Site Alternative Name in cases where a program is in the practice of using a different version of the Facility Site's name than the one recorded in Facility Site Name, such as a modified version required by their own regulations and procedures. This name may be one that the program will continue to use regularly, or it may be phased out in preference of the agency's Facility Site Name as the transition to an agency-wide standard is made.
Examples:	Amoco-Fairfield Deaver River Texaco Spill Site
Attribute:	Environmental Interest Type
Definition:	The reason that the agency is interested in a Facility Site.
Usage:	Most of these reasons are caused by regulations (either Federal or State), but there are also some that are due to new initiatives that may not have been implemented by rule (e.g., compliance assistance). The permitted values for this attribute, along with a description of each, should be recorded in a separate look-up table.
Examples:	TVSOURCE (Title V Permittee) P2PLAN (Pollution Prevention Planning)
Attribute:	Environmental Interest Start Date
Definition:	The date (mm-dd-yyyy) on which the Facility Site became of interest to the program for the reason described by the Environmental Interest Type.
Usage:	The meaning of this date will differ according to the Environmental Interest Type and should be reflected in the Environmental Interest Start Date Qualifier attribute.
Examples:	11-02-1994 06/12/1996
Attribute:	Environmental Interest End Date
Definition:	The date (mm-dd-yyyy) on which the Facility Site ceased to be of interest to the program for the reason described by the Environmental Interest Type.
Usage:	As with the Environmental Interest Start Date, the meaning of this date is determined by the nature of the Environmental Interest Type and should be reflected in the Environmental Interest End Date Qualifier attribute.
Examples:	12/30/1982 04-06-1997

Attribute: Environmental Interest Start Date Qualifier
 Definition: The qualifier that specifies the meaning of the date being used as an approximation for the environmental interest start date.

Usage: This information qualifies or describes the origin and quality of the Environmental Interest Start Date.

Examples: Date of First Report
 Date of Permit Issued

Attribute: Environmental Interest End Date Qualifier
 Definition: The qualifier that specifies the meaning of the date being used as an approximation for the environmental interest end date.

Usage: This information qualifies or describes the origin and quality of the Environmental Interest End Date.

Examples: Date Operations Ended
 Date Permit Ended

Attribute: Environmental Interest Information System Abbreviated Name
 Definition: The abbreviated name that represents the name of an information management system for an environmental program.

Usage: This information coupled with the Environmental Information System Identification Number allows more detailed information (i.e. permit, compliance) to be linked to the Facility Site. The permitted values for this attribute, along with a description of each, should be recorded in a separate look-up table.

Examples: SWIMS (State Water Information Management System)
 PDS (Permit Data System)

Entity: Alternative Identification

Definition: An identifier by which a particular program that has an interest in the Facility Site knows of it. The identifier may be a permit number issued by the program, an identification number for the Facility Site used in the program's files or database, a historic Facility Site name, or a Federal database identification number (or any other identifier a state finds useful to include).

Usage: Many of these alternative identification numbers may be needed to support links to older or legacy systems. Over time, these forms of identification may become unnecessary as agency-wide use of the State Facility Site Identifier becomes standardized.

Relationships between Entities:

- An Alternative Identification is always an alias used by a program with an Environmental Interest in the Facility Site

Attribute: Alternative ID Number
 Definition: An identification number by which the Facility Site is referred to in an existing system.

Examples: PAD001296729
 HW35132

Attribute: Current System Code
 Definition: An acronym or abbreviation for the system that contains additional information about the Facility Site. This should be the system to which the Alternative ID Number applies.

Usage: The permitted values for this attribute, along with a description of each, should be recorded in a separate look-up table.

Examples: RCRIS, NJPDES

Attribute: Facility Site Historic Name

Definition: An historic identification name by which the Facility Site is referred to in an existing system

Examples: Bob’s Photo

Attribute Formats

<i>Entity</i>	<i>Attribute</i>	<i>Length</i>	<i>Type</i>	<i>Requirement</i>
---------------	------------------	---------------	-------------	--------------------

Facility Site

	State Facility Identifier* ¹⁴	12	Alphanumeric	Mandatory
	Facility Site Federal Identifier* ¹⁵	12	Alphanumeric	Optional
	Facility Site Name*	80	Alphanumeric	Mandatory

Environmental Interest

	Environmental Information System Identification Number*	30	Alphanumeric	Mandatory
	Environmental Interest Site Alternative Name	60	Alphanumeric	Optional
	Environmental Interest Type*	60	Alphanumeric	Mandatory
	Environmental Interest Start Date*	N/A	Date	Mandatory
	Environmental Interest End Date*	N/A	Date	Optional
	Environmental Interest Start Date Qualifier*	50	Alphanumeric	Optional
	Environmental Interest End Date Qualifier*	50	Alphanumeric	Optional
	Environmental Information System Abbreviated Name*	50	Alphanumeric	Optional

Alternative Identification

	Alternative ID Number	20	Alphanumeric	Optional
	Current System Code	10	Alphanumeric	Optional
	Historic Facility Site Name	80	Alphanumeric	Optional

¹⁴ * indicates that the attribute exists in the State/EPA Facility Identification Data Standard.

¹⁵ Equivalent to the “Facility Registry Identifier” attribute in the State/EPA Facility Identification Data Standard.

GEOGRAPHIC LOCATION

Entity: *Geographic Location*

Definition: The physical location of a Facility Site, which may be described using up to three types of information: physical address and description, geographic coordinates, and geopolitical areas.

Usage: Provides the physical location of the Facility Site. Attributes of Geographic Location are highly interrelated.

Relationships between Entities:

- A geographic location is always associated with one Facility Site

Attribute: Location Identifier

Definition: A unique number assigned to each Geographic Location by the system.

Usage: This number is used primarily for data administration purposes, such as supporting electronic data interchange. This number is not necessarily useful to the data user.

Example: 014209353

Attribute: Location Address Line 1

Definition: The first line of the geographic (physical) address of the Facility Site. This may not be the same as its mailing address.

Example: 225 Huntington St

Attribute: Location Address Line 2

Definition: The second line of the geographic (physical) address of the Facility Site. This may not be the same as its mailing address.

Example: Huntington Towers Bldg, Ste 101

Attribute: Locality Name

Definition: The name of the city or town in which the Facility Site is located.

Examples: Boston
Walla Walla

Attribute: Location State USPS Code

Definition: The two-letter USPS abbreviation for the state in which the Site is located.

Usage: Each Facility Site must be physically located in the State. Although this attribute may seem unnecessary for a *state system*, it is included for completeness of the locational address and, optionally, as a crosscheck to ensure that only State Facility Sites are being added (avoiding possible confusion with mailing addresses). This code corresponds to the State Name attribute and should be recorded in a separate look-up table.

Examples: MA
NJ

Attribute: State Name

Definition: The name of a principal administrative subdivision of the United States, Canada, or Mexico.

Usage: Each Facility Site must be physically located in the State. Although this attribute may seem unnecessary for a *state system*, it is included for completeness of the locational address and, optionally, as a crosscheck to ensure that only State Facility Sites are being added (avoiding possible confusion with mailing

	addresses). This name corresponds to the Location State USPS Code attribute and should be recorded in a separate look-up table.
Examples:	Massachusetts New Jersey
Attribute:	Location ZIP Code/International Postal Code
Definition:	The 5-digit ZIP code of the address where the Facility Site is located.
Examples:	12343 98765
Attribute:	Location ZIP Code +4 Extension
Definition:	The +4 extension to the 5-digit ZIP code for the place where the Facility Site is located. Also known as the Geographic Segment Code.
Examples:	5404 9496
Attribute:	Location FIPS County Code
Definition:	The Federal Information Processing Standard (FIPS) code for the county where the Facility Site is located.
Usage:	This code may be entered directly into the system or derived from the location address or geographic coordinates. This code corresponds to the County Name attribute and should be recorded in a separate look-up table.
Examples:	5404 9496
Attribute:	County Name
Definition:	The name of the U.S. county or county equivalent in which the Facility Site is physically located.
Usage:	This name corresponds to the Location FIPS County Code and should be recorded in a separate look-up table.
Examples:	King Snohomish
Attribute:	Country Name
Definition:	The name that represents the primary geopolitical unit of the world in which the Facility Site is physically located. <i>Default:</i> United States.
Examples:	Canada Mexico
Attribute:	Tribal Land Name
Definition:	The name of an American Indian or Alaskan native area where the Facility Site is located.
Examples:	Pueblo of Nambe Fort Belknap Indian Community
Attribute:	Supplemental Location Text
Definition:	A brief explanation of where the Facility Site is located. This may include navigational directions and/or more descriptive information about the location of the Facility Site.
Example:	Take road #4, then turn right on #42. The building is at the end of road #42 on the left-hand side.
Attribute:	Latitude Measure
Definition:	The decimal representation of the latitude of the Facility Site location.
Example:	+45.123456

Attribute:	Longitude Measure
Definition:	The decimal representation of the longitude of the Facility Site location.
Example:	-116.123456
Attribute:	Geometric Type Code
Definition:	The code that represents the geometric entity represented by one geographic location.
Usage:	This code may be entered directly into the system or derived from the location address or geographic coordinates. This code corresponds to the Geometric Type Name and should be recorded in a separate look-up table.
Examples:	003 (Area) 005 (Route)
Attribute:	Geometric Type Name
Definition:	The name that identifies the geometric entity represented by one geographic location.
Usage:	These names correspond to the Geometric Type Code attribute and should be recorded in a separate look-up table.
Examples:	Area Route
Attribute:	Horizontal Collection Method Code
Definition:	The code that represents the method used to determine the latitude and longitude coordinates for a point on earth.
Usage:	This code corresponds to the Horizontal Collection Method Text attribute and should be recorded in a separate look-up table.
Examples:	01 02
Attribute:	Horizontal Collection Method Text
Definition:	The technique used to collect the coordinates of a Facility Site.
Usage:	This text corresponds to the Horizontal Collection Method Code attribute and should be recorded in a separate look-up table.
Examples:	Address Matching - Block Face Zip Code Centroid
Attribute:	Horizontal Accuracy Measure
Definition:	The accuracy value as a range (+/-) of the latitude and/or longitude.
Usage:	The value of the least accurate coordinate measurement should be used. In many cases Horizontal Accuracy Measure may be derived from the Horizontal Collection Method Code/Text information. The permitted values for this attribute, along with a description of each, should be recorded in a separate look-up table.
Example:	+/-10000
Attribute:	Horizontal Accuracy Measure Unit
Definition:	The units of the Horizontal Accuracy Measure value.
Usage:	The permitted values for this attribute, along with a description of each, should be recorded in a separate look-up table.
Examples:	Degrees Minutes Meters

Attribute:	Horizontal Reference Datum Code
Definition:	The code that represents the reference datum used in determining latitude and longitude coordinates.
Usage:	This code corresponds to the Horizontal Reference Datum Name attribute and should be recorded in a separate look-up table.
Examples:	NAD27 NAD83
Attribute:	Horizontal Reference Datum Name
Definition:	The standard shape or spheroid to which the surface of the earth has been orthographically transferred to depict the latitude/longitude.
Usage:	This name corresponds to the Horizontal Reference Datum Code attribute and should be recorded in a separate look-up table.
Examples:	North American Datum 1927 North American Datum 1983 ('91 Adj.)
Attribute:	Reference Point Code
Definition:	The code that represents the place for which geographic coordinates were established.
Usage:	This code corresponds to the Reference Point Text attribute and should be recorded in a separate look-up table.
Examples:	02 10
Attribute:	Reference Point Text
Definition:	The text that identifies the geographic position of the point within the Facility Site at which the coordinates of the Facility Site were determined.
Usage:	This text corresponds to the Reference Point Code attribute and should be recorded in a separate look-up table.
Examples:	NE Corner of Land Parcel Unknown
Attribute:	Reference Point Geographic Position Description
Definition:	A more detailed description of the geographic point at which the coordinates of the Site were determined.
Example:	Terminus of north discharge pipe
Attribute:	Source Map Scale Number
Definition:	The number that represents the proportional distance on the ground for one unit of measure on the map or photo.
Usage:	This attribute should be mandatory for all horizontal data collection methods except for methods using Global Positioning Systems (GPS). The permitted values for this attribute, along with a description of each, should be recorded in a separate look-up table.
Examples:	01 (1:10,000) 02 (1:100,000)
Attribute:	Congressional District Number
Definition:	The number of the Congressional District in which the Facility Site is located.
Usage:	The permitted values for this attribute, along with a description of each, should be recorded in a separate look-up table.
Example:	01 (Congressional District No. 1)

Attribute: State Legislative District Number
Definition: The number of the State Legislative District in which the Facility Site is located.
Usage: The permitted values for this attribute, along with a description of each, should be recorded in a separate look-up table.
Example: 12 (Legislative District No. 12)

Attribute Formats

<i>Entity</i>	<i>Attribute</i>	<i>Length</i>	<i>Type</i>	<i>Requirement</i> ¹⁶
Geographic Location				
	Location Address Line 1*	25	Alphanumeric	Mandatory (X)
	Location Address Line 2* ¹⁷	25	Alphanumeric	Optional
	Locality Name*	60	Alphanumeric	Mandatory (X)
	Location State USPS Code	2	Alphanumeric	Mandatory (X)
	State Name*	35	Alphanumeric	Mandatory
	Location ZIP Code/International Postal Code*	5	Alphanumeric	Optional
	Location FIPS County Code*	3	Alphanumeric	Optional
	County Name*	35	Alphanumeric	Optional
	Country Name*	44	Alphanumeric	Optional
	Tribal Land Name*	52	Alphanumeric	Optional
	Supplemental Location Text*	50	Alphanumeric	Optional
	Latitude Measure*	9	Alphanumeric	¹⁸ Mandatory (Z)
	Longitude Measure*	10	Alphanumeric	¹⁹ Mandatory (Z)
	Geometric Type Code*	3	Alphanumeric	Optional
	Geometric Type Name*	6	Alphanumeric	Optional
	Horizontal Collection Method Code*	3	Alphanumeric	Optional
	Horizontal Collection Method*	60	Alphanumeric	Optional
	Horizontal Accuracy Measure*	6	Alphanumeric	Optional
	Horizontal Accuracy Measure Units	6	Alphanumeric	Optional
	Horizontal Reference Datum Code*	3	Alphanumeric	Optional
	Horizontal Reference Datum Name*	7	Alphanumeric	Optional
	Reference Point Code*	3	Alphanumeric	Optional
	Reference Point Text*	2	Alphanumeric	Optional
	Reference Point Geographic Position Description	50	Alphanumeric	Optional
	Source Map Scale Number*	10	Alphanumeric	Optional
	Congressional District Number	2	Alphanumeric	Optional
	State Legislative District Number	2	Alphanumeric	Optional

INDUSTRY CLASSIFICATION***Entity: Industry Classification***

¹⁶ A Site must have a Geographic Location. To be meaningful, the Geographic Location information must include at least one of the following types of information: physical address (for which the necessary attributes are marked with an 'X'); and/or a coordinate (for which the necessary attributes are marked with a 'Z').

¹⁷ "Location Address Line 1 and Line 2" attributes are equivalent to the State/EPA Facility Identification Data Standard's "Location Address" attribute.

¹⁸ This numeric field must use the 'mask': +/-12.123456 (i.e., two digits before the decimal and six digits after the decimal)

¹⁹ This numeric field must use the 'mask': +/-123.123456 (i.e., three digits before the decimal and six digits after the decimal)

Definition: The type of industry or business conducted at the Facility Site.
Usage: This provides a general description of the type of activities conducted at the Facility Site.

SIC and NAICS codes are used to describe the type of industry. Because various activities may be performed at a Facility Site, more than one SIC or NAICS code may apply to any one Facility Site. Industry Classification may also be associated with specific Environmental Interests.

Relationships between Entities:

- An Industry Classification sometimes describes the type of activities performed at one Facility Site
- An Industry Classification is sometimes associated with a specific Environmental Interest

Attribute: Standard Industrial Classification (SIC) Code
Definition: A 2-, 3-, or 4-digit Standard Industrial Classification code that is drawn from a hierarchical classification system of industries based on the products they produce and/or services they perform.
Usage: The permitted values for this code, along with a description of each, should be recorded in a separate look-up table.
Examples: 20 (Food and Kindred Products)
 201 (Meat Products)
 2011 (Meat and Packing Plants)

Attribute: SIC Primary Indicator
Definition: The name that indicates whether the associated SIC Code represents the primary activity occurring at the Facility Site.
Usage: The permitted values for this code, along with a description of each, should be recorded in a separate look-up table. If a primary SIC Code cannot be identified the 'Unknown' value should be recorded.
Examples: Primary
 Secondary
 Unknown

Attribute: U.S. National Industry Classification System (NAICS) Code
Definition: The code that represents a subdivision of an industry that accommodates user needs in the United States.
Usage: The permitted values for this code, along with a description of each, should be recorded in a separate look-up table.
Examples: 325520 (Adhesive Manufacturing)
 337211 (Wood Office Furniture Manufacturing)

Attribute: NAICS Primary Indicator
Definition: The name that indicates whether the associated NAICS Code represents the primary activity at the Facility Site.
Usage: The permitted values for this code, along with a description of each, should be recorded in a separate look-up table. If a primary NAICS code cannot be identified the 'Unknown' value should be recorded.
Examples: Primary
 Secondary
 Unknown

Attribute Formats

<i>Entity</i>	<i>Attribute</i>	<i>Length</i>	<i>Type</i>	<i>Requirement</i>
Industry Classification				
	Standard Industrial Classification (SIC) Code*	4	Alphanumeric	Mandatory
	SIC Primary Indicator*	10	Alphanumeric	Optional
	North American U.S. National Industry Classification System (NAICS) Code*	6	Alphanumeric	Optional
	NAICS Primary Indicator*	10	Alphanumeric	Optional

AFFILIATION

Entity: Affiliation

Definition: A relationship between a Facility Site and an Individual or Organization.

Usage: The Affiliation defines the roles that external Organizations and Individuals play at a Facility Site. Such relationships between Organizations, Individuals, and a Facility Site provide the agency with the names, phone numbers, and mailing addresses that the agency needs in order to communicate with a Facility Site. Some Affiliations represent current relationships with the people affiliated with the Facility Site, and some represent past relationships.

Although each Affiliation specifies only one Organization or Individual, the same Organization or Individual may be associated many times with the same Facility Site through other Affiliations. This feature accommodates situations where the same Organization or Individual plays many roles at a Facility Site (e.g., owner, operator, and site contact). Furthermore, any one Organization or Individual may be affiliated with many Facility Sites. This is particularly important as it allows the agency to determine all the roles an Organization or Individual plays with Facility Sites in the state.

The Affiliation Type (see attribute below) will likely dictate how an Affiliation should be recorded given the agency's rules for programmatic sharing of, and historical tracking of, Affiliation. More specifically, the Affiliation Type will determine whether an Affiliation is one shared by the entire agency or is specific to one environmental interest. The legal owner of a Facility Site should only need to be recorded once, regardless of each Environmental Interest in the Facility Site, whereas the Facility Site contact may be different for each environmental interest in the Facility Site, and need to be recorded several times.

In addition, Affiliation Type will determine whether the agency must track the history of previous affiliations. When the responsible party for a Facility Site no longer plays that role, the agency must record that the Affiliation has ended and the date on which it ended, but cannot remove any reference to it. However, when a Facility Site Contact has left the site, the agency does not need to track the contact's historical Affiliation with the Facility Site.

These rules may be enforced by formal working procedures or may be enforced by the system. If they are enforced by the system, a separate look-up table is needed for each Affiliation Type.

Relationships between Entities:

- An Affiliation always involves one Facility Site
- An Affiliation sometimes involves an Individual
- An Affiliation sometimes involves an Organization
- An Affiliation sometimes involves a Mailing Address
- An Affiliation is sometimes associated with a specific Environmental Interest

Attribute: Affiliation Type
Definition: The relationship between the affiliated Individual and/or Organization and the Facility Site.
Usage: The permitted values for this attribute, along with a description of each, should be recorded in a separate look-up table.
Examples: LGLOWNER (Legal Owner)
ATTORNEY (Attorney)

Attribute: Phone Number
Definition: The primary phone number of a Facility Site affiliate, either an Individual or Organization. The phone number should include a three-digit area code and a seven-digit phone number.
Usage: If both an Individual and an Organization are listed, the Phone Number is generally associated with the Individual. The length of the phone number should be long enough to allow international phone numbers to be entered.
Examples: (602) 555- 4444
01144 1224 762347

Attribute: Phone Number Extension
Definition: The extension number of the primary phone number for a Facility Site affiliate.
Usage: If both an Individual and an Organization are listed, the extension number is generally associated with the Individual.
Examples: 429
8125

Attribute: E-mail Address
Definition: The electronic mail address of a Facility Site affiliate.
Usage: If both an Individual and an Organization are listed, the electronic mail address is generally associated with the Individual.
Examples: ray.beard@enviro.edu
dolf_hinden@fairfieldcorp.com

Attribute: Fax Number
Definition: The phone number to which a facsimile can be sent to a Facility Site affiliate. The phone number should include a three-digit area code and a seven-digit phone number.
Usage: If both an Individual and an Organization are listed, the Fax Number is generally associated with the Individual. The length of the phone number should be long enough to allow international phone numbers to be entered.
Examples: (510) 792-3067
46599 2357 331251

Attribute: Alternate Phone Number
Definition: An alternate phone number of a Facility Site affiliate, either an Individual or Organization.
Usage: If both an Individual and an Organization are listed, the Alternate Phone Number is generally associated with the Individual. The length of the phone number should be long enough to allow international phone numbers to be entered.
Examples: (541) 555-6789
01214 8729 233701

Attribute: Alternate Phone Number Extension
Definition: The extension number of the alternate phone number for a Facility Site affiliate.

Usage: If both an Individual and an Organization are listed, the Alternate Phone Number Extension is generally associated with the Individual.

Examples: 555378
3405

Attribute: Affiliation Start Date

Definition: The date (mm-dd-yyyy) on which the affiliation between the Facility Site and Individual and/or Organization began.

Usage: The meaning of “active” is determined by an agency’s policies regarding different affiliation types. For example, the Start Date of a “legal owner” must be exactly when the Facility Site came under that new ownership; whereas the Start Date of a “Facility Site contact” might be when the agency first heard about the new contact.

Examples: 01-23-1997
10/25/1973

Attribute: Affiliation End Date

Definition: The date (mm-dd-yyyy) on which the affiliation between the Facility Site and Individual and/or Organization ended.

Usage: Similar to Affiliation Start Date, the meaning of this date is determined by agency policy regarding need for historical information.

Examples: 11/09/1988
06-23-1997

Entity: Individual

Definition: A person that is of interest to the agency because of that person’s interactions with one or more Facility Sites (e.g., a person who is the contact for three Facility Sites and is the responsible party for another).

Usage: It is important for the agency to avoid duplication of Individuals, so that the agency can answer questions such as, “What other sites is this Individual involved with?”

Relationships between Entities:

- An Individual is always a person involved in one or more Affiliations.

Attribute: Individual Identifier

Definition: The number assigned to the Individual by the system that uniquely identifies the Individual.

Usage: This number is used primarily for data administration purposes, such as supporting electronic data interchange. This number is not necessarily useful to the data user.

Examples: 1234
34-8521

Attribute: Last Name

Definition: The last name of the Individual who is affiliated with a Facility Site.

Examples: Smith
Heraldson

Attribute: First Name

Definition: The first name of the Individual who is affiliated with a Facility Site.

Examples: John

	Jane
Attribute:	Middle Initial
Definition:	The middle initial of the Individual who is affiliated with a Facility Site.
Usage:	In some cases, the middle initial is the only way the agency can differentiate one person from another. ²⁰
Examples:	W T
Attribute:	Individual Title Text
Definition:	The Individual's job title.
Usage:	This information can be useful when the Individual who acts as the contact person for an Organization that is affiliated to a Facility Site leaves the Organization; when the agency attempts to contact that Individual, the job title may help the Organization determine an alternative Individual for the agency to speak with.
Examples:	Environmental Manager EHS Administrator

Entity: Mailing Address

Definition:	The standard address used to send mail to an Individual or Organization through the US Postal Service.
Usage:	A mailing address is primarily used as the means for the agency to formally correspond with an affiliated Individual or Organization. A Mailing Address is not intended to allow the agency to physically locate an affiliate (or even a Facility Site), but only to allow them to mail something to the affiliate.

For international mailing addresses (which typically represent only a very small percentage of cases), the country must be specified. Where the format of the address does not match the attributes below (e.g., instead of city, a district or other jurisdiction type), such information should be included in the Address Line 1 and Address Line 2 attribute fields. This would include mailing codes for international addresses—ZIP codes are only required for US addresses.

Relationships between Entities:

- A Mailing Address is always the address for one or more Affiliation

Attribute:	Mailing Address Identifier
Definition:	The number assigned to the Mailing Address by the system that uniquely identifies the Mailing Address.
Usage:	This number is used primarily for data administration purposes, such as supporting electronic data interchange. This number is not necessarily useful to the data user.
Examples:	294 2358905

²⁰ An alternative to the Middle Initial would be the full (middle) name. However, it is very often the case that when people write their name on a form they may only include the initial without spelling out the full middle name. If the system had only a middle name field then there would be a good possibility that the same person would be duplicated because of inconsistencies across multiple data collection mechanisms (i.e. 'Brian Ralph Smith' may or may not be the same as 'Brian R. Smith'). To limit this issue only the initial is included in the model.

Attribute:	Mailing Address Line 1
Definition:	The first line of the mailing address of an Organization and/or Individual.
Example:	12045 E 39 th
Attribute:	Mailing Address Line 2
Definition:	The second line of the mailing address of an Organization and/or Individual.
Usage:	This attribute can also store international addresses when necessary, since the attributes are free format.
Example:	World Trade Center, Ste 4318
Attribute:	Supplemental Address Text
Definition:	The text that provides additional information to facilitate the delivery of a mail piece, including building name, secondary units, and mail stop or local box numbers not serviced by the U.S. Postal Service.
Example:	Building name is First Tower, go to floor 14, suite 1402, mail stop 14001.
Attribute:	Mailing Address City Name
Definition:	The name of the city or town in the mailing address of an Organization and/or Individual
Usage:	The city name is the full name with no abbreviations, unless the abbreviation is inherent in the formal name (e.g., Pt. Angeles). If the address is in an unincorporated area, the Mailing Address City Name is the name assigned by the post office for mail delivery.
Examples:	Los Angeles Philadelphia
Attribute:	Mailing Address State USPS Code
Definition:	The two-letter postal abbreviation for the state in the mailing address of an Organization and/or Individual
Usage:	This code corresponds to the Mailing Address State Name attribute and should be recorded in a separate look-up table.
Examples:	PA ME
Attribute:	Mailing Address State Name
Definition:	The name of the state where the mail is delivered.
Usage:	This name corresponds to the Mailing Address State USPS Code attribute and should be recorded in a separate look-up table.
Examples:	Pennsylvania Maine
Attribute:	Mailing Address ZIP Code/International Postal Code
Definition:	The 5-digit postal ZIP code for the mailing address of an Organization and/or Individual.
Examples:	97202 46015
Attribute:	Mailing Address ZIP Code +4 Extension
Definition:	The +4 extension to the 5-digit ZIP code for the mailing address of an Organization and/or Individual. Also known as the Geographic Segment Code.
Examples:	5355 1203

Attribute:	Mailing Address Country Code
Definition:	The code for the country in the mailing address of an Organization and/or Individual.
Usage:	The default entry for this field should be US. This code corresponds to the Mailing Address Country Name attribute and should be recorded in a separate look-up table.
Examples:	UK FR
Attribute:	Mailing Address Country Name
Definition:	The name of the country where the addressee is located.
Usage:	The default entry for this field should be the United States. This name corresponds to the Mailing Address Country Code and should be recorded in a separate look-up table.
Examples:	United Kingdom France

Entity: Organization

Definition:	A company, government body, or similar entity that has some responsibility or role at one or more Facility Sites (e.g., an Organization that operates five Facility Sites and is the owner of two of them).
Usage:	It is important for the agency to avoid duplication of Organizations, so that the agency can answer questions such as, “What other sites is this Organization involved with?”

Relationships between Entities:

- An Organization is always involved in one or more Affiliations.

Attribute:	Organization Identifier
Definition:	The number assigned to the Organization by the system that uniquely identifies the Organization.
Usage:	This number is used primarily for data administration purposes, such as supporting electronic data interchange. This number is not necessarily useful to the data user.
Examples:	4523 327543245
Attribute:	Organization DUNS Number
Definition:	The Data Universal Numbering System (DUNS) number assigned by Dun and Bradstreet to identify unique business establishments.
Examples:	59-100-5301 (A F Ross Athol Ttd) 76-017-9650 (William Baxter Limited)
Attribute:	Organization Formal Name
Definition:	The formal name of the Organization.
Usage:	The Organization Name should be the name as it is determined by a State’s individual business rules. These rules would define how the name is identified and how it should be formatted (e.g., abbreviations).
Examples:	Bradley-Noble Geotech Co. Masterpower Inc

Attribute Formats

<i>Entity</i>	<i>Attribute</i>	<i>Length</i>	<i>Type</i>	<i>Requirement</i>
Affiliation				
	Affiliation Type*	40	Alphanumeric	Mandatory
	Phone Number	15	Alphanumeric	Optional
	Email Address	40	Alphanumeric	Optional
	Fax Number	15	Alphanumeric	Optional
	Alternate Phone Number	15	Alphanumeric	Optional
	Affiliation Start Date	N/A	Date	Mandatory
	Affiliation End Date	N/A	Date	Optional
Individual				
	Last Name*	30	Alphanumeric	Mandatory
	First Name*	30	Alphanumeric	Mandatory
	Middle Initial* ²¹	1	Alphanumeric	Optional
	Individual Title Text*	40	Alphanumeric	Optional
Mailing Address ²²				
	Mailing Address Line 1*	25	Alphanumeric	Mandatory
	Mailing Address Line 2* ²³	25	Alphanumeric	Optional
	Supplemental Address Text*	50	Alphanumeric	Optional
	Mailing Address USPS State Code	2	Alphanumeric	Mandatory
	Mailing Address State Name*	35	Alphanumeric	Mandatory
	Mailing Address City Name*	30	Alphanumeric	Mandatory
	Mailing Address ZIP Code/International Postal Code*	10	Alphanumeric	Mandatory
	Mailing Address Country Code	2	Alphanumeric	Mandatory
	Mailing Address Country Name*	44	Alphanumeric	Mandatory
Organization				
	Organization DUNS Number*	9	Alphanumeric	Optional
	Organization Formal Name*	80	Alphanumeric	Mandatory

²¹ “Individual Last Name, First Name, and Middle Initial” attributes are equivalent to the State/EPA Facility Identification Data Standard’s “Individual Full Name” attribute.

²² The required attributes for a Mailing Address shown in the table represent those for a US postal address. For international mailing addresses, the Country Code (populated with a non-US code) would override the required US address attributes, such that only the Address Identifier, Address Lines 1 and 2, and Country Code would be applied.

²³ “Mailing Address Line 1 and Line 2” attributes are equivalent to the State EPA Facility Identification Data Standard’s Mailing Address attribute.

APPENDIX B: GENERALIZED ERROR SCENARIOS

During interviews with states for FITS II, common scenarios highlighting a variety of system and business rule issues were identified. The error scenarios detailed below illustrate how the model, architectural approaches and workflows might behave under several realistic problems. They are intended as illustrations only; any resemblance to a single state is coincidental. Each scenario begins with a generic description of a user and an error. It then describes how that error and its resolution might work under each of the example architectures. The scenarios use the term “system” to refer the application the user considers his primary working application. When originally conceived, these scenarios were designed to formally illustrate very specifically how the architectures differed. As they were developed, however, it became clear that the response of any system to the errors indicated here would be dependent on the complex interaction of many factors including a) the volatility of the agency’s Facility Site data, b) the architecture, c) the enterprise data model, and d) the business practices followed by data entry staff. There are literally hundreds of relevant permutations of these, far too many to cover usefully. Instead, these scenarios proved to be useful for two different reasons:

- They provide examples that managers can use to undertake a reality check of their designs – especially the interaction between business practices and system logic.
- They illustrate the perspective of the harried program user in system maintenance. While most systems are conceived and designed with the objective of providing holistic integrated perspectives on Facility Sites, most of the staff responsible for the data have very specific and immediate concerns focused on one programmatic or other area.

SCENARIO ONE—DUPLICATE CANDIDATES FROM SYSTEM OVERSIGHT

Larry is permit writer for the Agency He has just received a stack of what are supposed to be brand new permit applications for a new stormwater program and must review and enter the applications into the system before he can go on vacation. Larry is aware of something about “the integrated system” and “the business rules” for data entry but he is new, and busy. Larry quickly enters the new permits as “new Facility Sites,” by passing the reminder screen that all new entries should be checked against the agency masterfile before entry. This is a new program, from small businesses, so Larry reasons that they are probably not in the file yet. He has more than enough crosschecking and verification to do on the permit data itself. Besides, once they are entered, he can run his report, deliver it to his boss and go on vacation.

SCENARIO ONE—ISSUES

Larry was only partly right. Two of the Facility Sites also had NPDES permits and were already in the master file. Duplicate Facility Site records now exist for these two Facility Sites.

SCENARIO TWO—DUPLICATE CANDIDATES FROM SYSTEM MISUSE

Renee runs a special hazardous waste permitting program. She is an expert Excel user and has created her own sophisticated spreadsheet of the 300 Facility Sites she and her staff work with. All of these Facility Sites have other permits with the agency. Until yesterday she did not take very seriously the agency’s “integration initiative,” but the new agency policy that all major Facility Site information assets must use the new system took her by surprise. Renee quickly discovers that almost all of her Facility Sites are in the system but some seem to have new names, and some have contact and address information different from what she is used to. Several staff (not yet trained on the system) have already complained that they cannot find their Facility Sites and that the systems reports are not as familiar as her Excel macros. With a big mailing coming up, Renee and her staff are in a rush. Some of the staff begin to use the system by updating the

masterfile and adding their own information in the environmental interest areas (and learning more about their Facility Sites). Other users become frustrated, especially for several of their individual large Facility Sites that seem to have been chopped into pieces. The staff quickly discovers, however, that they can simply create “new” Facility Sites that have just the information they need, as they are used to seeing it. They can run their reports, and they can also avoid whatever the “reconciliation” process is that seems to be distracting some of the other staff.

SCENARIO TWO—ISSUES

Of the 300 Facility Sites Renee’s branch works with, 290 were in the masterfile. Of these, 25 were “duplicated” by staff who wanted to preserve the format of their programmatic information. Several of these can be identified by the SWP (Special Waste Program) text one user appended to the names.

SCENARIOS ONE AND TWO ANALYSIS—DUPLICATE CANDIDATES

In the scenarios above, duplicate records are introduced into the agency system, either through an oversight or misuse of the system (or both). In a warehouse architecture users only enter data into their program system and are not presented with any pro-active indication that there may be a duplicate present. In fact there is no duplicate in the program system, the duplicate only occurs when the next update merges this data with the enterprise data. The warehouse architecture may support the strongest sense that the problem is not “owned” by the program office but is instead an artifact of the extraction process (which may be correct). The program system is otherwise entirely independent of the enterprise system. This independence is most significant when the program offices, for whatever reason, are unresponsive to these flagged errors. The errors will persist until corrected in the program system, backed out of the enterprise system, and then re-refreshed using the corrected data. It is unlikely any non-program staff (i.e., the stewards of the central system) would or could make these changes directly to the program data through the program applications. As an alternative, a state may encourage users to first check the central system before entering new Facility Sites, but the system may have no way to enforce the business rules. This may also require users to switch applications in the middle of their work.

Under hybrid and centralized architectures, the users are interacting directly with the enterprise Facility Site data as they enter new Facility Sites. Searches for pre-existing Facility Sites take place in the same system used to enter them. While at least one state system requires a search before creating a record, most rely on a business rule and training to encourage the user to perform this search. System managers even report that some users “spoof” the required search, or, as in scenario two, change the name just enough to avoid triggering the duplicate checking system. Once in the system, the duplicate must be detected by the agency’s quality assurance scans and corrected through the resolution/reconciliation process.

Several states with hybrid or centralized architecture have also taken an organizational design approach to this problem by restricting the creation of new Facility Sites to a small group of staff equipped, trained and held accountable for maintaining the integrity of the data. This obviously represents a significant renegotiation of the traditional program independence in data management.

In the hybrid and central architectures, there is both a stronger perceptual and more direct system connection between the program and enterprise data. Program staff routinely interact with enterprise data and routinely see their data in this larger context. In addition, these architectures provide a more direct ability to check for, communicate, and ultimately correct inconsistencies than does the warehouse system because they provide a bi-directional connection between the

program and enterprise data. Staff use the same system to enter their programmatic data and to investigate and reconcile problems. In addition, the integrity of the enterprise and program Facility Site data can be actively maintained by the central staff. If so authorized, central staff can also write directly to the program data.

Although the warehouse architecture may have to wait for the next update cycle to begin its scan, the architectures are quite similar in the workflow of the *error detection* process. Most states implement both automated and manual reviews of all additions and updates using some combination of name, address and/or location data to detect duplicates, and some logical and data standard validation checks. Errors that cannot be directly corrected are flagged and listed, triggering a work process where staff investigate and correct the problem. Here the differences in architecture have more indirect but important implications for the ultimate success of resolution. For the remainder of this discussion it is presumed that some program data must be corrected to “fix” the example problem.

Two things have to happen before a problem can be fixed, a program staff has to be aware, and has to care. While an email or other more active alert may suffice to flag errors to the designated staff, the sense of ownership and motivation to fix the data must come from elsewhere. By itself, the warehouse architecture may allow the impression the problem is not “owned” by the program office but is instead an artifact of the extraction process (which may be correct) or some other problem “over there”. Aside from these extracts the program system is otherwise entirely independent of the enterprise system. This independence is most significant in the common situation where the program offices, for whatever reason, are unresponsive to these flagged errors. Program staff need some motivation or incentive to fix these errors.. Until fixed in the program system, and then re-refreshed using the corrected data, the error will remain in the enterprise system. In most cases non-program staff (i.e. the stewards of the central system) cannot make these fixes directly to the program data but must rely on program staff to do so.

Finally, the last scenarios illustrate issues that are largely independent of architecture. Instead, they demonstrate the importance of the data model (i.e., its capture of relevant metadata) and the organizational capacity to investigate, resolve, and jointly implement fixes for the inevitable problems. Scenario three profiles an agency conflict resulting in immature or absent update and maintenance protocols. In scenario four, each program office is doing its best to follow the reconciliation procedures but because of the conflict of references the data do not agree. This scenario re-enforces the finding that reconciliation procedures do not seek to find THE absolute right answer but only one that is mutually agreeable, valid and reproducible enough for the system to function.

SCENARIO THREE—LACK OF UPDATE AND MAINTENANCE

On her return from a national meeting, the agency’s secretary has just announced an information initiative. Staff is directed to “integrate the agency’s information and make it available on the web.” A team of staff members is formed and given a small budget; one of the staff has just returned from a data-warehousing workshop. The team immediately begins to debate the approach they should take and splits into two factions. One (who had read the FITS document) argued that, despite the history of difficulty in cross-agency efforts, the program offices must be involved from the beginning, especially in determining the business rules that would be used for establishing and updating the new agency information base, even if it meant not meeting the Secretary’s target. The other faction, armed with the latest database connectivity and web-server tools, argues that involving the program offices will take too long, and that once the system is mounted and introduced, the update and maintenance rules can be figured out. They also argue

that IT staff will be able to interface with the program offices as needed to resolve the Facility Site data and investigate any reconciliation issues. This faction wins. Four months later, the secretary announces the new system with much fanfare. Curious program staff find that for the first time, they are able to view cross-program data and print maps of their Facility Sites. There are many places where the data in the system does not match their data, but these can be ignored or fixed manually. After three months, however, a more serious problem arises, the data is out of date. None of the program offices have responded to the now atrophied team's request for updates of the data, and the programmer who performed the original population has left the agency. As the data grows more and more out of date, staff become less and less inclined to use it, and even less responsive to the team's requests for updates. After one year, funding for system is redistributed and the warehouse is taken down.

SCENARIO FOUR—THE RIGHT PAGE

Tom is part of his program's data integration team. The agency has just completed the first phase of implementation for its Facility Site profiler system. Users have been moderately happy with the system, but the rules for who can change whose data have yet to be fully resolved, and this has caused some friction and hesitancy to rely on the data. The water office had been one of the first to bring its Facility Sites into the system; this included some quite old data on many of its Facility Sites. As the new watershed office began to use this data, they began to update the agency locational data for several Facility Sites in target areas. Normally the agency follows the rule that the user with the most recent and most accurate data can update the location of a Facility Site, but this has started to cause problems when the office of water uses this updated data to update its maps. For an unknown reason, the Facility Sites don't match up, yet everybody's data seems right to them. This causes several of the staff to begin to refuse to use the central data and revert to their historical local records for fear that somebody else will mess up their data. This also results in several Facility Site records being duplicated by staff so that they can assert their "primary" location. Upon investigation, a GIS staff discovers that the offices have been using a different reference datum, and the mismatch between program offices was an artifact of this different metadata. With some effort, staff negotiates a new business rule that makes the GIS metadata more prominent in the update process, establishes rules for locational date priorities, and provides links to historical locational data.

APPENDIX C: STATE EPA FACILITY IDENTIFICATION STANDARD

STATE EPA INFORMATION MANAGEMENT WORKGROUP TRANSMITTAL LETTER

To: The State EPA Information Management Workgroup
From: Facility Identification Action Team co-chairs
Date: July 19, 1999
RE: The Proposed Facility Identification Data Standard

With this memorandum the Facility Identification Action Team transmits to the State EPA Information Management Workgroup (the Workgroup), the *Proposed Facility Identification Data Standard*. In addition to the standard itself (see attached), this memorandum communicates the Action Team's findings and recommendations about the standard, and its use in the context of the State-EPA data relationship.

These findings and recommendations are identical to those reviewed in the *Facility Identification Action Team Workgroup Decision Memo*, also forwarded to the Workgroup under separate cover.

The Action Team Product: Proposed Facility Identification Data Standard

The standard itself is very simple, it describes the data elements used to uniquely identify a Facility Site and differentiate it from other Facility Sites. The standard contains an introduction, which provides context to its application, and defines several groups of data elements by listing their name, definition and format. The standard also describes the relationship of these groups to the main Facility Site group. The groups are:

- Facility Site
- Geographic Coordinates
- Mailing Address
- Affiliation
- Organization
- Individual
- Environmental Interest
- Standard Industrial Classification [incorporated by reference]
- North American Industry Classification [incorporated by reference]

The Action Team held extensive, substantive discussion and debates on most of these groups and their data elements. The final proposed standard represents a significant evolution from the Action Team's starting point. Elements were added, deleted and re-defined at the suggestion of State and EPA Action Team members. There is a general sense that this standard is as good as current knowledge and experience can make it, and that it is an excellent start. All participants expect the standard to be revised over time as EPA and States gain experience using the standard and exchanging data.

Findings

The standard in the context of the State-EPA data relationship

Expert staff from States and EPA have labored over this proposal and believe it to be sound. The Action Team's deliberations included policy and implementation issues of the standard and of how EPA and or States would pursue facility identification. Because many of these issues had never been engaged from a joint State-EPA perspective, they were an appropriate and consistent element of Action Team deliberations. As the standard itself solidified, the Action Team began to re-engage these issues and forge the resolutions contained in the standard's introduction. The language used in the standard itself (shown in italics below) is necessarily terse. Therefore these resolutions are highlighted and elaborated on below, they are:

1) *The standard provides guidance to those developing systems to manage facility data and to data trading partners who wish to exchange facility identification information.* In the context of the State EPA data relationship, the standard's most important use will be in supporting the coordination and exchange of facility identification data where States and EPA have agreed to do so.

2) *The standard does not establish new or modify existing data collections or reporting requirements.* EPA and States are mutually obligated by a complex web of data agreements, requirements, and historical *de facto* standards. These obligations are embodied in statutes, regulations, delegation agreements, and other individual State-EPA agreements. The standard itself does not change these.

EPA and States do aspire, through efforts of the Workgroup, to rationalize their mutual data flows by developing standards (beyond facility identification) for environmental data that each agree to use. The Data Standards Strategy Action Team is recommending a broader strategy for the joint development and use of data standards.

3) *The standard is not intended to represent a minimum nor a maximum set of data that an agency should collect, manage or exchange to meet its facility identification business needs.* States and EPA have, or are developing, facility identification processes. These are supported by agency systems, collections and data standards. In many instances the data elements of the standard exceed what is needed for a given collection or exchange; in others, they are only a start.

4) (As a corollary to #2 and #3) The standard does not prescribe individual elements as [mandatory] or [optional]. *The standard allows trading partners and data managers to establish rules about the specific elements they need to manage, as their business needs dictate. It also avoids any confusion that might result from elements prescribed as "mandatory" being interpreted by others as establishing or modifying a data collection or reporting requirement.*

Recommendations

The Action Team recommends the Workgroup take the following actions:

1. Accept the facility standard forwarded by the Action Team.
2. Forward the standard to the appropriate body for States and EPA endorsement.
3. Actively encourage and monitor use of the standard in evolving State and EPA data flows.

We look forward to presenting and discussing these recommendations with you later this month.

Acknowledgement

The chairs would also like to take this opportunity to thank the Action Team members for their tireless participation in what proved to be a difficult, long but ultimately successful joint endeavor. We hope this success is a harbinger of future joint successes in use of the standard itself, and ultimately in improving data we use to protect and preserve our environment

[signed] Lynn Singleton, Washington, DOE William A. Sonntag, USEPA

State-EPA Information Management Workgroup
Facility Identification Action Team, Co- Chairs

**State-EPA Information Management Workgroup
Facility Identification Action Team Members:**

EPA Members

William Sonntag, Co-Chair	Office of Reinvention
David Hindin	OECA
Ken Blumberg	Region 1
Dave Levy	Office of Solid Waste
Leonard Fitch	Office of Water
John Sullivan	One Stop/ Office of Reinvention
Ed Lillis, Chuck Isbell	Office of Air
Jack Diskin	Office of Pollution Prevention and Toxic Substances
George Bonina	Office of Information Resource Management

State Members

Lynn Singleton, Co-Chair	Washington DOE
Cathy Wagenfer	Maryland DE
Audrey Surette	Indiana DEM
Nadeem Siddiqui	Alaska DEC
Ann Bidwell	Minnesota PCA
Daniel Miller	Arizona DEQ
Gail Jackson	Pennsylvania DEP

Support Contractors

Louis Sweeny	Ross & Associates
Dalroy Ward	SAIC, Systems Development Center

APPROVED STATE EPA FACILITY IDENTIFICATION STANDARD*FINAL DRAFT - JULY 19, 1999*

This standard describes the data elements used to uniquely identify a Facility Site and differentiate it from other Facility Sites. The standard was developed by a state/EPA Action Team (team) chartered by the ECOS-State-EPA Information Management Workgroup. The standard provides guidance to those developing systems to manage facility data and to data trading partners who wish to exchange facility identification information

“Facility” and “Site” are terms that have been defined differently in various environmental regulations and programs. This standard relies on the common English definitions of facility and site, in order to accommodate multiple technical definitions of both terms.

The data elements are organized into groups (i.e., Facility Site, Geographic Coordinates, Affiliation, Organization, Individual, Mailing Address, Environmental Interest, Standard Industrial Classification, and North American Industry Classification). Each group represents a different thing of significance, related to the identification of a Facility Site, about which information needs to be known. For each group, a definition is provided, and a specification of its relationships to the other groups. Facility Site is the central group, and all other groups are related to it either directly or indirectly.

This standard does not establish new or modify existing data collections, reporting requirements or system development requirements. This standard addresses data elements most relevant to the identification of a Facility Site. The team also agreed that the data elements which describe basic information about the Facility Site and the environmental interests associated with the Facility Site are always necessary. The standard allows trading partners and data managers to establish rules about the specific elements they need to manage, as their business needs dictate. It also avoids any confusion that might result from elements prescribed as mandatory being interpreted by others as establishing or modifying a data collection or reporting requirement.

The team recognizes that agencies (both state and EPA) will designate elements and relationships as mandatory under specific circumstances and that implementation of this standard in existing or future data exchanges will require the development and/or negotiation of any programmatic changes or exchange-specific rules. The standard is not intended to represent a minimum nor a maximum set of data that an agency should collect, manage or exchange to meet its facility identification business needs. These are implementation issues outside the scope of the standard itself.

A definition and format is given for each data element. The format provides the maximum length of the data element and the data type. Data types include alphanumeric (A), number (N), and Date. For several data elements, allowable values are also included. In most cases, the allowable values shown are provided as examples to illustrate the intended use of their respective element. These are noted as examples in the text. In a few cases, however, where the standard makes use of a reference set defined elsewhere (e.g. FIPS codes) or where the set of allowable values is small and stable, the values shown or referenced represent the initial permitted code set for their respective elements.

DATA ELEMENT NAME	DATA ELEMENT DEFINITION	FORMAT
<p>Facility Site <i>Definition:</i> Basic identification information for a Facility Site, including the facility registry identifier, geographic address, and geopolitical descriptors. <i>Relationships:</i> Each Facility Site may be involved with one or more Affiliation(s). Each Facility Site may be classified by one or more Standard Industrial Classification(s). Each Facility Site may be classified by one or more North American Industry Classification(s). Each Facility Site may be geographically located by one or more Geographic Coordinates. Each Facility Site must be monitored by one or more Environmental Interest(s).</p>		
Facility Registry Identifier	The identification number assigned by the EPA Facility Registry System to uniquely identify a Facility Site.	A(12)
State Facility Identifier	The unique identification number used by a state to identify a Facility Site.	A(12)
Facility Site Name	The public or commercial name of a Facility Site (i.e., the full name that commonly appears on invoices, signs, or other business documents, or as assigned by the state when the name is ambiguous).	A(80)
Location Address	The address that describes the physical (geographic) location of the front door or main entrance of a Facility Site, including urban-style street address or rural address.	A(50)
Supplemental Location Text	The text that provides additional information about a place, including a building name with its secondary unit and number, an industrial park name, an installation name or descriptive text where no formal address is available.	A(50)
Locality Name	The name of the city, town, village or other locality, when identifiable, within whose boundaries (the majority of) the Facility Site is located. This is not always the same as the city used for USPS mail delivery. <i>Allowable Values: (examples)</i> “None” is an allowable value. The code set found in the current FIPS 55 Guideline: <i>Codes for Named Populated Places, Primary County Divisions, and Other Locational Entities of the United States, Puerto Rico, and the Outlying Areas.</i> The URL is: http://www.itl.nist.gov/div897/pubs/fip55-3.htm .	A(60)
County and State FIPS Code	The code that represents the county or county equivalent and the state or state equivalent of the United States. <i>Allowable Values:</i> All codes for counties and county equivalents of all states of	A(5)

DATA ELEMENT NAME	DATA ELEMENT DEFINITION	FORMAT
	the U.S. as well as U.S. territories and possessions found in the current FIPS publication 6-4, <i>Counties and Equivalent Entities of the United States, Its Possessions, and Associated Areas</i> . <i>Remarks:</i> The first 2-digits of the code represent the state; the last 3-digits represent the county. For example, 09001 represents Fairfield County (001), Connecticut (09).	
County Name	The name of the U.S. county or county equivalent in which the Facility Site is physically located.	A(35)
State Name	The name of a principal administrative subdivision of the United States, Canada, or Mexico.	A(35)
Country Name	The name that represents a primary geopolitical unit of the world. <i>Default:</i> United States	A(44)
Location ZIP Code/ International Postal Code	The combination of the 5-digit Zone Improvement Plan (ZIP) code and the four-digit extension code (if available) that represents the geographic segment that is a subunit of the ZIP Code, assigned by the U.S. Postal Service to a geographic location; or the postal zone specific to the country, other than the U.S., where the Facility Site is located.	A(14)
Tribal Land Name	The name of an American Indian or Alaskan native area where the Facility Site is located.	A(52)

Geographic Coordinates

Definition:

A geographic point, or set of points, defined by latitude and longitude coordinates used to locate a Facility Site, usually the front door or centroid, including the associated method, accuracy, and description data.

Relationships:

Each Geographic Coordinates occurrence must geographically locate one and only one Facility Site occurrence.

Remarks:

This group is included by reference to the *EPA Latitude/Longitude Data Standard*; only mandatory data elements are shown. For an example allowable values list, see the Environmental Data Registry for EPA's list (URL: <http://www.epa.gov/edr/>). There can be multiple Geographic Coordinates associated with a Facility Site, however, each instance of a Geographic Coordinate can only be associated with one Facility Site occurrence.

Latitude Measure	The measure of the angular distance on a meridian north or south of the equator.	A(6) - A(10) DD.dddddd
Longitude Measure	The measure of the angular distance on a meridian east or west of the prime meridian.	A(7) - A(11) DDD.dddd d

DATA ELEMENT NAME	DATA ELEMENT DEFINITION	FORMAT
Horizontal Accuracy Measure	The measure of the accuracy (in meters) of the latitude and longitude coordinates.	A(6) in meters
Geometric Type (Textual Data or Code Data acceptable)		
Code	The code that represents the geometric entity represented by one point or a sequence of latitude and longitude points.	A(3)
Name	The name that identifies the geometric entity represented by one point or a sequence of latitude and longitude points.	A(6)
Horizontal Collection Method (Textual Data or Code Data acceptable)		
Code	The code that represents the method used to determine the latitude and longitude coordinates for a point on the earth.	A(3)
Text	The text that describes the method used to determine the latitude and longitude coordinates for a point on the earth.	A(60)
Horizontal Reference Datum (Textual Data or Code Data acceptable)		
Code	The code that represents the reference datum used in determining latitude and longitude coordinates.	A(3)
Name	The name that describes the reference datum used in determining latitude and longitude coordinates.	A(7)
Reference Point (Textual Data or Code Data acceptable)		
Code	The code that represents the place for which geographic coordinates were established.	A(3)
Text	The text that identifies the place for which geographic coordinates were established.	A(50)
10Map Scale Number	The number that represents the proportional distance on the ground for one unit of measure on the map or photo. <i>Remarks:</i> Mandatory for all horizontal data collection methods except for methods using Global Positioning System (GPS).	A(10)

Affiliation

Definition:

The relationship between a Facility Site and an organization and/or an individual person.

Relationships:

Each Affiliation occurrence must be established with one and only one Facility Site occurrence.

Each Affiliation occurrence may involve one and only one Organization occurrence.

Each Affiliation occurrence may involve one and only one Individual occurrence.

Each Affiliation occurrence may receive mail at one and only one Mailing Address occurrence.

Remarks:

This doesn't imply that the affiliation must exist, but; if an affiliation exists, it must be associated with one and only one Facility Site. An Organization may participate with multiple Facility Sites. Each type of

DATA ELEMENT NAME	DATA ELEMENT DEFINITION	FORMAT
<p>affiliation can exist more than once for a Facility Site; thus there can be two occurrences of the Legally Responsible Entity affiliation type with a Facility Site.</p>		
<p>Affiliation Type</p>	<p>The name that describes the capacity or function that an organization or individual serves for a Facility Site. <i>Allowable Values (examples):</i> <u>Organization</u> Legally Responsible Entity Legal Operator Waste Treater Waste Handler Land Owner Parent Corporation</p>	<p>A (40)</p>
<p>Organization <i>Definition:</i> A company, government body, or other type of organization that has some responsibility or role at the Facility Site. <i>Relationships:</i> Each Organization must be involved with one or more Affiliation(s).</p>		
<p>Organization Formal Name</p>	<p>The legal, formal name of an organization that is affiliated with the Facility Site.</p>	<p>A(80)</p>
<p>Organization DUNS Number</p>	<p>The Data Universal Numbering System (DUNS) number assigned by Dun and Bradstreet to identify unique business establishments</p>	<p>A(9)</p>
<p>Individual <i>Definition:</i> An individual person who has some responsibility or role at the Facility Site. <i>Relationships:</i> Each Individual must be involved with one or more Affiliation(s).</p>		
<p>Individual Full Name</p>	<p>The complete name of a person, including first name, middle name or initial, and surname.</p>	<p>A (70)</p>
<p>Individual Title Text</p>	<p>The title held by a person in an organization.</p>	<p>A(40)</p>
<p>Mailing Address <i>Definition:</i> The standard address used to send mail to an individual or organization affiliated with the Facility Site. <i>Relationships:</i> Each Mailing Address must be the delivery point for one or more Affiliation(s).</p>		
<p>Mailing Address</p>	<p>The exact address where a mail piece is intended to be delivered, including urban-style street address, rural route, and PO Box.</p>	<p>A(50)</p>
<p>Supplemental Address Text</p>	<p>The text that provides additional information to facilitate the delivery of a mail piece, including building name, secondary</p>	<p>A(50)</p>

DATA ELEMENT NAME	DATA ELEMENT DEFINITION	FORMAT
	units, and mail stop or local box numbers not serviced by the U.S. Postal Service.	
Mailing Address City Name	The name of the city, town, or village where the mail is delivered.	A(30)
Mailing Address State Name	The name of the state where mail is delivered.	A(35)
Mailing Address Country Name	The name of the country where the addressee is located. <i>Default: United States</i>	A(44)
Mailing Address ZIP Code/International Postal Code	The combination of the 5-digit Zone Improvement Plan (ZIP) code and the four-digit extension code (if available) that represents the geographic segment that is a subunit of the ZIP Code, assigned by the U.S. Postal Service to a geographic location to facilitate mail delivery; or the postal zone specific to the country, other than the U.S., where the mail is delivered.	A(14)

Environmental Interest

Definition:

The environmental permits and regulatory programs that apply to the Facility Site.

Relationships:

Each Environmental Interest occurrence must apply to one and only one Facility Site occurrence.

Environmental Interest Type	<p>The environmental permit or regulatory program that applies to the Facility Site.</p> <p><i>Allowable Values: (examples)</i></p> <table border="0"> <thead> <tr> <th><u>Value</u></th> <th><u>Meaning</u></th> </tr> </thead> <tbody> <tr> <td>TRI Report</td> <td>Toxic Release Inventory Report</td> </tr> <tr> <td>NPDES Major</td> <td>Clean Water Act NPDES Major</td> </tr> <tr> <td>NPDES Minor</td> <td>Clean Water Act NPDES Minor</td> </tr> <tr> <td>NPDES General Permit</td> <td>Clean Water Act NPDES General Permit</td> </tr> <tr> <td>NPDES Stormwater Permit</td> <td>Clean Water Act NPDES Stormwater Permit</td> </tr> <tr> <td>NPDES Other Permit</td> <td>Clean Water Act NPDES Other Permit</td> </tr> <tr> <td>Pretreatment SIU</td> <td>Clean Water Act Pretreatment Significant Industrial User</td> </tr> <tr> <td>SS Major</td> <td>Clean Air Act Stationary Source Major</td> </tr> <tr> <td>SS Synthetic Minor</td> <td>Clean Air Act Stationary Source Synthetic Minor</td> </tr> <tr> <td>SS NESHAP Minor</td> <td>Clean Air Act Stationary Source NESHAP Minor</td> </tr> <tr> <td>SS Other Minor</td> <td>Clean Air Act Stationary Source Other Minor</td> </tr> <tr> <td>RMP Facility</td> <td>Clean Air Act RMP Facility</td> </tr> <tr> <td>TSD</td> <td>Hazardous Waste Handler -</td> </tr> </tbody> </table>	<u>Value</u>	<u>Meaning</u>	TRI Report	Toxic Release Inventory Report	NPDES Major	Clean Water Act NPDES Major	NPDES Minor	Clean Water Act NPDES Minor	NPDES General Permit	Clean Water Act NPDES General Permit	NPDES Stormwater Permit	Clean Water Act NPDES Stormwater Permit	NPDES Other Permit	Clean Water Act NPDES Other Permit	Pretreatment SIU	Clean Water Act Pretreatment Significant Industrial User	SS Major	Clean Air Act Stationary Source Major	SS Synthetic Minor	Clean Air Act Stationary Source Synthetic Minor	SS NESHAP Minor	Clean Air Act Stationary Source NESHAP Minor	SS Other Minor	Clean Air Act Stationary Source Other Minor	RMP Facility	Clean Air Act RMP Facility	TSD	Hazardous Waste Handler -	A(60)
<u>Value</u>	<u>Meaning</u>																													
TRI Report	Toxic Release Inventory Report																													
NPDES Major	Clean Water Act NPDES Major																													
NPDES Minor	Clean Water Act NPDES Minor																													
NPDES General Permit	Clean Water Act NPDES General Permit																													
NPDES Stormwater Permit	Clean Water Act NPDES Stormwater Permit																													
NPDES Other Permit	Clean Water Act NPDES Other Permit																													
Pretreatment SIU	Clean Water Act Pretreatment Significant Industrial User																													
SS Major	Clean Air Act Stationary Source Major																													
SS Synthetic Minor	Clean Air Act Stationary Source Synthetic Minor																													
SS NESHAP Minor	Clean Air Act Stationary Source NESHAP Minor																													
SS Other Minor	Clean Air Act Stationary Source Other Minor																													
RMP Facility	Clean Air Act RMP Facility																													
TSD	Hazardous Waste Handler -																													

DATA ELEMENT NAME	DATA ELEMENT DEFINITION	FORMAT
	<p>Treatment, Storage, Disposal (TSD)</p> <p>LQG (EPA Defined) Hazardous Waste Handler - Large Quantity Generator (LQG), as defined by EPA</p> <p>LQG (State Defined) Hazardous Waste Handler - Large Quantity Generator (LQG), as defined by State</p> <p>SQG Hazardous Waste Handler - Small Quantity Generator (SQG)</p> <p>Spill Control Plan Oil Pollution Act Spill Control Plan</p> <p>UIC Underground Injection Control Well (UIC)</p> <p>UST Underground Storage Tank (UST)</p> <p><i>Remarks:</i> This list will be expanded as necessary to include values for additional interests.</p>	
Environmental Interest Start Date	Date the agency became interested in the Facility Site for a particular environmental interest type.	Date
Environmental Interest End Date	Date the agency ceased to be interested in the Facility Site for a particular environmental interest type.	Date
Environmental Interest Start Date Qualifier	<p>The qualifier that specifies the meaning of the date being used as an approximation for the environmental interest start date.</p> <p><i>Allowable Values: (examples)</i></p> <p>Date of First Report</p> <p>Date Operations Commenced</p> <p>Date of Permit Application</p> <p>Date Permit Issued</p> <p>Date Monitoring Started</p>	A(50)
Environmental Interest End Date Qualifier	<p>The qualifier that specifies the meaning of the date being used as an approximation for the environmental interest end date.</p> <p><i>Allowable Values: (examples)</i></p> <p>Date of last report</p> <p>Date Permit Ended</p> <p>Date Operations Ended</p>	A(50)
Environmental Information System Abbreviated Name	The abbreviated name that represents the name of an information management system for an environmental program.	A(15)
Environmental Information System Identification Number	The identification number, such as the permit number, assigned by an information management system that represents a Facility Site, waste site, operable unit, or other	A(30)

DATA ELEMENT NAME	DATA ELEMENT DEFINITION	FORMAT								
	feature tracked by that Environmental Information System.									
Standard Industrial Classification										
<i>Definition:</i>										
The Standard Industrial Classification (SIC), or type of business activity, occurring at the Facility Site.										
<i>Relationships:</i>										
Each Standard Industrial Classification occurrence must classify one and only one Facility Site occurrence.										
<i>Remarks:</i>										
This group is included by reference to the <i>SIC/NAICS Data Standard</i> .										
Standard Industrial Classification Code	The code that represents the economic activity of a company (4-digits).	A(4)								
SIC Primary Indicator	The name that indicates whether the associated SIC Code represents the primary activity occurring at the Facility Site. <i>Allowable Values:</i> <table border="0" data-bbox="638 777 1364 1050"> <tr> <td><u>Value</u></td> <td><u>Meaning</u></td> </tr> <tr> <td>Primary</td> <td>The SIC Code represents the primary activity occurring at the Facility Site.</td> </tr> <tr> <td>Secondary</td> <td>The SIC Code represents a secondary activity occurring at the Facility Site.</td> </tr> <tr> <td>Unknown</td> <td>It is not known whether the SIC Code represents the primary or secondary activity at the Facility Site.</td> </tr> </table>	<u>Value</u>	<u>Meaning</u>	Primary	The SIC Code represents the primary activity occurring at the Facility Site.	Secondary	The SIC Code represents a secondary activity occurring at the Facility Site.	Unknown	It is not known whether the SIC Code represents the primary or secondary activity at the Facility Site.	A(10)
<u>Value</u>	<u>Meaning</u>									
Primary	The SIC Code represents the primary activity occurring at the Facility Site.									
Secondary	The SIC Code represents a secondary activity occurring at the Facility Site.									
Unknown	It is not known whether the SIC Code represents the primary or secondary activity at the Facility Site.									
North American Industry Classification										
The North American Industry Classification System (NAICS) code, or type of industrial activity, occurring at the Facility Site.										
<i>Relationships:</i>										
Each North American Industry Classification must classify one and only one Facility Site.										
<i>Remarks:</i>										
This group is included by reference to the <i>SIC/NAICS Data Standard</i> .										
North American Industry Classification System Industry Code	The code that represents a subdivision of an industry that accommodates user needs in the United States (6-digits).	A(6)								
NAICS Primary Indicator	The name that indicates whether the associated NAICS Code represents the primary activity occurring at the Facility Site. <i>Allowable Values:</i> <table border="0" data-bbox="638 1564 1364 1827"> <tr> <td><u>Value</u></td> <td><u>Meaning</u></td> </tr> <tr> <td>Primary</td> <td>The NAICS Code represents the primary activity occurring at the Facility Site.</td> </tr> <tr> <td>Secondary</td> <td>The NAICS Code represents a secondary activity occurring at the Facility Site.</td> </tr> <tr> <td>Unknown</td> <td>It is not known whether the NAICS Code represents the primary or secondary activity at the Facility Site.</td> </tr> </table>	<u>Value</u>	<u>Meaning</u>	Primary	The NAICS Code represents the primary activity occurring at the Facility Site.	Secondary	The NAICS Code represents a secondary activity occurring at the Facility Site.	Unknown	It is not known whether the NAICS Code represents the primary or secondary activity at the Facility Site.	A(10)
<u>Value</u>	<u>Meaning</u>									
Primary	The NAICS Code represents the primary activity occurring at the Facility Site.									
Secondary	The NAICS Code represents a secondary activity occurring at the Facility Site.									
Unknown	It is not known whether the NAICS Code represents the primary or secondary activity at the Facility Site.									