

# Development of the EIIP Emission Inventory Data Model for Purposes of Facilitating Data Transfer Between the States and U.S. EPA

**John Slade**

Pennsylvania Department of Environmental Resources,  
P.O. Box 8468, 400 Market Street, Harrisburg, PA 17105-8468

**William Benjey\***

Atmospheric Sciences Modeling Division, Air Resources Laboratory,  
National Oceanic and Atmospheric Administration, Research Triangle Park, NC 27711

**Andrew Blackard and Michael Pring**

Eastern Research Group, Inc.,  
1600 Perimeter Park, Suite 300, Morrisville, NC 27560

## ABSTRACT

The Emission Inventory Improvement Program (EIIP) was formed with the cooperation of the State and Territorial Air Pollution Program Administrators/Association of Local Air Pollution Control Officials (STAPPA/ALAPCO) as a cooperative effort between State and local agencies, industry, and EPA to improve emission estimates and establish a National Standard for Emission Data Exchange. Emission data are needed for a variety of purposes, including regional air quality modeling, air quality planning, trends analysis, and for informing the public. The Data Management Committee (DMC) of the EIIP was charged with the goal of developing the data exchange standard, and undertook the task of developing a data model on which to base the data exchange format. The purpose of developing the data model was to establish a standardized set of data relationships which reflect physical reality and ensure accurate and consistent reporting of emissions and emission related data.

## INTRODUCTION

The Clean Air Act Amendments of 1990 (CAAA) require that many states collect emission inventory data as a basis for planning and demonstrating attainment with National Ambient Air Quality Standards (NAAQS). Emission inventories are a basic component of air quality modeling and other air pollution management analyses that are used to demonstrate that proposed air pollution control strategies are sufficient to attain the state and federal air quality standards.

Sharing of emission data is essential for urban and regional air quality modeling exercises that cross state boundaries, for regional and national regulatory impact and other policy analyses, for scientific research efforts, and for informing the public. In order to more easily share data among the emission inventory community, common formats for data sharing must be established. Such common formats can be based on an underlying data model which establishes the relationships of the data elements reflected within the common format.

The Emission Inventory Improvement Program (EIIP) Data Management Committee (DMC) has been formed through STAPPA/ALAPCO as a joint effort of state and local air agencies, industry, and the U.S. Environmental Protection Agency (EPA) to improve the way data/information is transferred and shared among users. The EIIP DMC is currently working to develop and coordinate recommendations for data transfer at the facility, state/local agency, and EPA levels. In addition, the DMC will develop recommended mechanisms for sharing emission inventory data and related information among all users.

\*On assignment to the National Exposure Research Laboratory, U.S. Environmental Protection Agency

The first step in achieving these goals was to develop the EIIP Phase I Data Model.<sup>1</sup> The focus of the Phase I Data Model is on the data needed for regional air quality modeling. Phases II and III are intended to add information needed for the air quality permitting process and other related emission inventory data needs of industry for reporting to state and local air pollution control agencies.

## **DATA MODEL DEVELOPMENT PROCESS**

The draft EIIP Phase I Data Model was developed from an emission inventory viewpoint by individuals knowledgeable in the origin and use of emission inventory data. The types of data needed by emission inventory users for modeling and developing emission inventories were identified from several sources, and then grouped and organized to construct the data model.

Initially, the EIIP Point, Area, Mobile, and Biogenic Sources Committees (source committees) submitted independent draft data models used to identify data elements needed in the EIIP Phase I Data Model as well as to show how the data elements related to each other. In addition to the source committee data models, a technical EPA memorandum (Seitz, 1995)<sup>2</sup> was used to identify the emission inventory data elements needed to support regional modeling. At the direction of the EIIP Steering Committee, these data elements were to be the focus of the EIIP Phase I Data Model.

Using these data, a preliminary data model was compiled, distributed back to the source committees for review, and discussed during subsequent teleconferences with each committee. The draft data model was also mapped against several existing state databases to identify any missing data elements and/or incorrect relationships among the data. This review process culminated in a workshop held in Research Triangle Park (RTP), North Carolina, in April 1996, and was attended by DMC members and state representatives involved in the mapping exercise. At this workshop, the DMC decided that it would be desirable to document the data model using a separate diagram for each source type (i.e., point, area and non-road mobile, on-road mobile, and biogenic). Each diagram would then represent the viewpoint of the user of that data (e.g., point sources) using familiar terminology.

The data model is currently being used to begin development of the EIIP data transfer format as well as an Oracle database at EPA for use in the EIIP Data Transfer Prototype project. Application of the data model in the context of the data transfer format and prototype database offered a final round of review and precipitated additional changes in the data model to ensure compatibility between the data model, data transfer format, and prototype database. Further information on these two applications may be found in the paper "Prototype Demonstration of Application of the EIIP Data Transfer Format" which is also being presented at this conference.

## **FINAL DATA MODEL**

A data model is a representation of data on paper. It is used to logically organize data and show how the data relate to each other. To categorize each individual piece of information requires specifying two organizational concepts--the entity and the attribute. The entity is the item (person, place, or thing) that the information describes. The attribute is a specific characteristic or description of an entity.

A data model is visually presented in the form of an entity-relationship diagram (ERD). The ERD shows each entity in the model and how they relate to one another.

For purposes of this paper, only the point source view of the EIIP Phase I Data Model is discussed in detail. Please refer to the final document "EIIP Phase I Data Model" for a

comprehensive description of the complete data model including the data element dictionary, data model coding schemes, examples, and further information on conventions of use for the data model.

Figure 1 presents an ERD showing the chief entities that provide the hierarchical framework for applying the EIIP Phase I Data Model to point sources. Figure 2 shows the complete ERD for point sources. This includes additional entities organized around the hierarchy of the "backbone" entities shown in Figure 1. Each of the "boxes" presented in the figure is an entity. A brief discussion of the intended use of each of these entities in the point source data model is provided below.

For the readers benefit, Figures 3, 4, and 5 are presented which show the interpretation of the EIIP Phase I Data Model for Area and Non-Road Mobile, On-Road Mobile, and Biogenic Sources, respectively. The primary "backbone" entities are discussed in hierarchical order as they appear on Figure 1, with the remaining "supporting" entities discussed in alphabetical order.

### **Site**

The Site entity identifies the facility or plant where the emissions are created and is at the top of the data model organizational hierarchy. The name, address, and contact information for the facility or plant is contained here. This entity also includes information such as the Standard Industrial Classification (SIC) Code for the facility, the number of employees, and the federal identification ID {e.g., Aerometric Information Retrieval System (AIRS) plant ID} for the facility. An abbreviated attribute listing for this entity is presented in Figure 6.

A Site contains one or more Emission Units, as indicated in Figure 2.

### **Emission Unit**

The Emission Unit entity contains information on the physical unit that creates the emissions. For instance, this could be a boiler or an incinerator. An abbreviated attribute listing for this entity is presented in Figure 7.

An Emission Unit has one or more Emission Processes as indicated in Figure 2.

### **Emission Process**

The Emission Process entity identifies the process(es) occurring at the Emission Unit that produces the emissions. For example, combustion of more than one type of fuel may occur in a single boiler. Each Emission Process is assigned a specific Source Classification Code (SCC). An abbreviated attribute listing for this entity is presented in Figure 8.

An Emission Process has one or more measures of activity in the Activity entity, as indicated in Figure 2.

### **Activity**

The Activity entity is used to represent activity information for a specific Emission Process. This would include information such as thruput or process rate. This information is divided into individual time periods with a measure of activity for each specific time period.

Each time period in the Activity entity will be linked to the emission of one or more pollutants in the Emissions entity. An abbreviated attribute listing for this entity is presented in Figure 9.

## **Emissions**

The Emissions entity represents the emission estimate. A separate instance of the Emissions entity is used to represent each pollutant. Therefore, each emission estimate is unique for a specific combination of Site, Emission Unit, Emission Process, Activity time period, and pollutant. The emission estimate may be designated as controlled or uncontrolled, and under a variety of operating scenarios (i.e., maximum, potential) and time periods (annual, daily, etc.,). An abbreviated attribute listing for this entity is presented in Figure 10.

## **Aggregate Controls as Applied**

The Aggregate Controls as Applied entity represents the combined, overall effect of the Control Strategy for an emissions estimate. This information is pollutant-specific and includes total capture efficiency, control efficiency, as well as rule effectiveness information. Also, when control data are provided, the corresponding emission estimate reflects the control efficiency indicated in the Aggregate Controls as Applied entity.

## **Control Strategy**

This entity identifies the complete collection of control measures that are applied to the linked Geographic Location, Site, Emission Unit, or Emission Process. The various rules, regulations, and corresponding emission limits that apply are named here.

## **Control Equipment**

The Control Equipment entity contains information for a specific control device. More than one control device may be linked to a single instance of the Aggregate Controls as Applied entity. The description and capacity of each individual control device is contained here. The connectivity of control devices to Emission Units, Emission Processes, stacks, or other control devices is indicated in the Path entity.

## **Control Equipment Characteristics**

This entity is related to the Control Equipment entity. The Control Equipment Characteristics entity is used to store the pollutant-specific attributes (percent capture efficiency, percent control efficiency) associated with the control device described in the Control Equipment entity.

## **Defined Areas**

A given Site may belong to more than one specially defined enforcement or study areas. The Defined Areas entity is intended to identify the organizational group that an area belongs to. The AIRS coding system is used for identifying specific nonattainment areas.

The Area Name attribute of the Defined Areas entity will be a free-form text attribute so that other specially defined areas may be specified if needed.

## **Emission Factors**

The Emission Factors entity is used to identify the specific emission factor used to calculate the emission estimate contained in the corresponding Emissions entity.

## **Emission Release Point**

This entity is used to identify the stack (or release point, such as a roof vent) the emissions are released from as well as to contain the stack's federal ID code for reporting. The entity is linked to the Path entity to indicate its connectivity to the Emission Process for which emissions are being reported. Alternatively, the connectivity of stacks to emission units or control devices may be indicated in the Path entity.

## **Geographic Coordinates**

The discrete coordinates for site locations, stack locations, and grids are located here. Appendix D of the EIIP Phase I Data Model document presents additional details on the use of this entity.

## **Geographic Location**

This entity contains geographic location information. This includes the following chief fields: country, state/province/territory, and county/parish/reservation. Other supplementary information such as air basin municipality, air quality control region, and study grids are also identified here. The discrete coordinates for points and grids are located in the Geographic Coordinates entity.

The location information in the Geographic Location entity is linked with other entities using a geographic unique identifiers (UID).

## **Meteorology**

This entity contains the meteorological data that were used to derive the supplied emission estimates. These data are linked to the Activity entity. Wind speed and temperature data are included in this entity.

## **Path**

The Path entity is used to show the relationships existing between any configuration of Emission Units, Emission Processes, Control Devices, and Emission Release Points (stacks). It may be applied to show the connectivity between one or more pairs of these physical pieces of equipment. The connectivity of Emission Units to Control Devices and Emission Release Points is indicated in Figure 2 by use of the Path entity. A more comprehensive discussion of application of the Path entity is provided in the EIIP Phase I Data Model document.

## **Process Growth Factors**

Annual growth factors for a given Emission Process are contained in the Process Growth Factor entity. This is typically the expected annual growth of the SCC assigned in the Emission Process entity. The Process Growth Factor entity includes the initial (base) year, a projected year, and a growth factor. The initial year corresponds to the year of the measure of emissions reported in the linked Emissions entity. The Process Growth Factor may be applied to this measure to estimate emissions for the projected year.

## **Stack Physical Parameters**

The type of stack is identified here. This entity also contains the physical dimensions of the stack and stack gas properties including temperature, velocity, and flow rate. The Geographic Coordinate entity contains stack location information and is linked through the Stack Geographic UID.

## **Transmittal Information**

The transmittal information entity is not a part of the hierarchy of entities that form the "backbone" of the EIIP Data Model. Instead it contains information about a particular data transmittal. This information includes the year of the inventory, the inventory's approval status, contact personnel, and other descriptive attributes.

## **SUMMARY**

The EIIP Data Model was developed through the collaboration of personnel from the EPA, state and local air pollution control agencies, industry, and support contractors. The intent of the data model is to define and organize the data elements (and relationships between data elements)

which are needed in air pollution studies, with particular emphasis on air quality modeling applications.

To date, the EIIP Data Model has been used in several applications to support the overall goals of the DMC. The primary application was to provide a structural basis for developing the standard EIIP Data Transfer format. The data model has also provided the basic blueprint for developing an Oracle database at EPA designed for use in the EIIP Data Transfer Prototype Demonstration, and for storing the EPA's National Emission Trends (NET) Inventory. In addition to the immediate goals of the DMC, the data model may also be used by States and local agencies considering revision or initial development of a relational database used for management of air pollution information.

As stated previously, the next steps for the evolution of the EIIP data model is expansion to include the information needed for permitting and compliance activities, as well as detailed facility level information required of industry by State and local agencies. It is not expected that these future phases would necessitate a re-design of the data model, but rather it would require additions to the current data model structure.

## **ACKNOWLEDGEMENTS**

This paper, and the Phase I Data Model Report upon which the paper is based, are the result of almost three years of conscientious and diligent effort by State and Federal members of the EIIP DMC, as well as the support contractors. Many thanks are also due to numerous other external reviewers and those who helped "map" the data model against State emission data models.

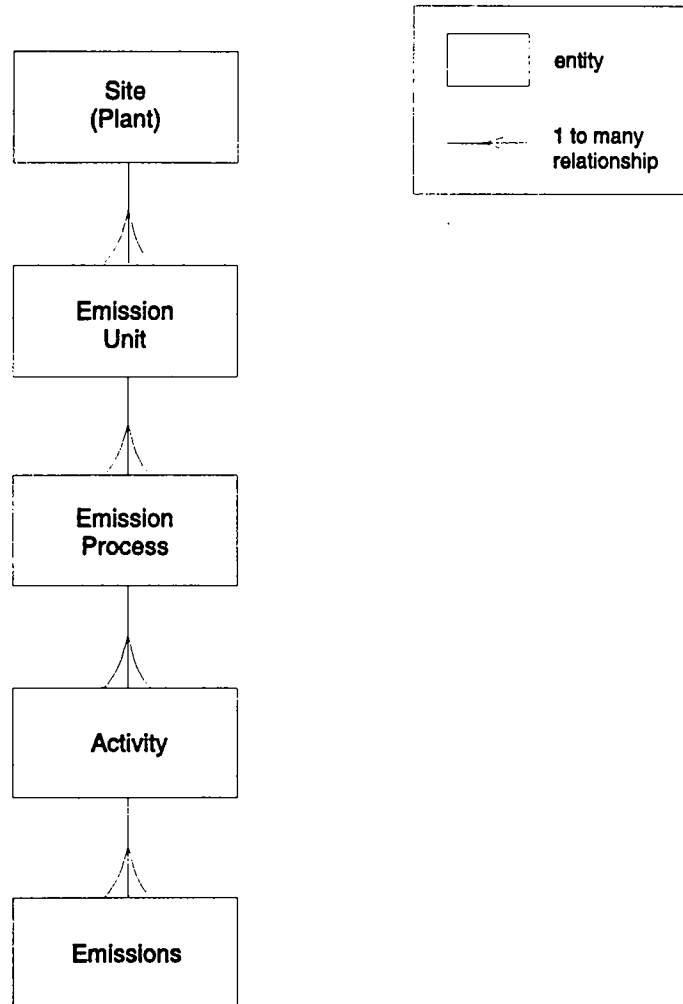
## **DISCLAIMER**

The information in this document has been funded wholly or in part by the United States Environmental Protection Agency. It has been subjected to Agency review and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

## **REFERENCES**

1. Emission Inventory Improvement Program, Data Management Committee, *EIIP Phase I Data Model*, Final Report, EIIP Document Series, Volume 7, U.S. Environmental Protection Agency, Research Triangle Park, NC, September 1997; EPA-454/R-97-004g (<http://www.epa.gov/oar/oaqps/eiip/dmchaps.html>)
2. Seitz, J.S. 1995. U.S. Environmental Protection Agency, Research Triangle Park, NC, *letter to State Air Directors*.

**Figure 1. Point Source Hierarchical Entities**



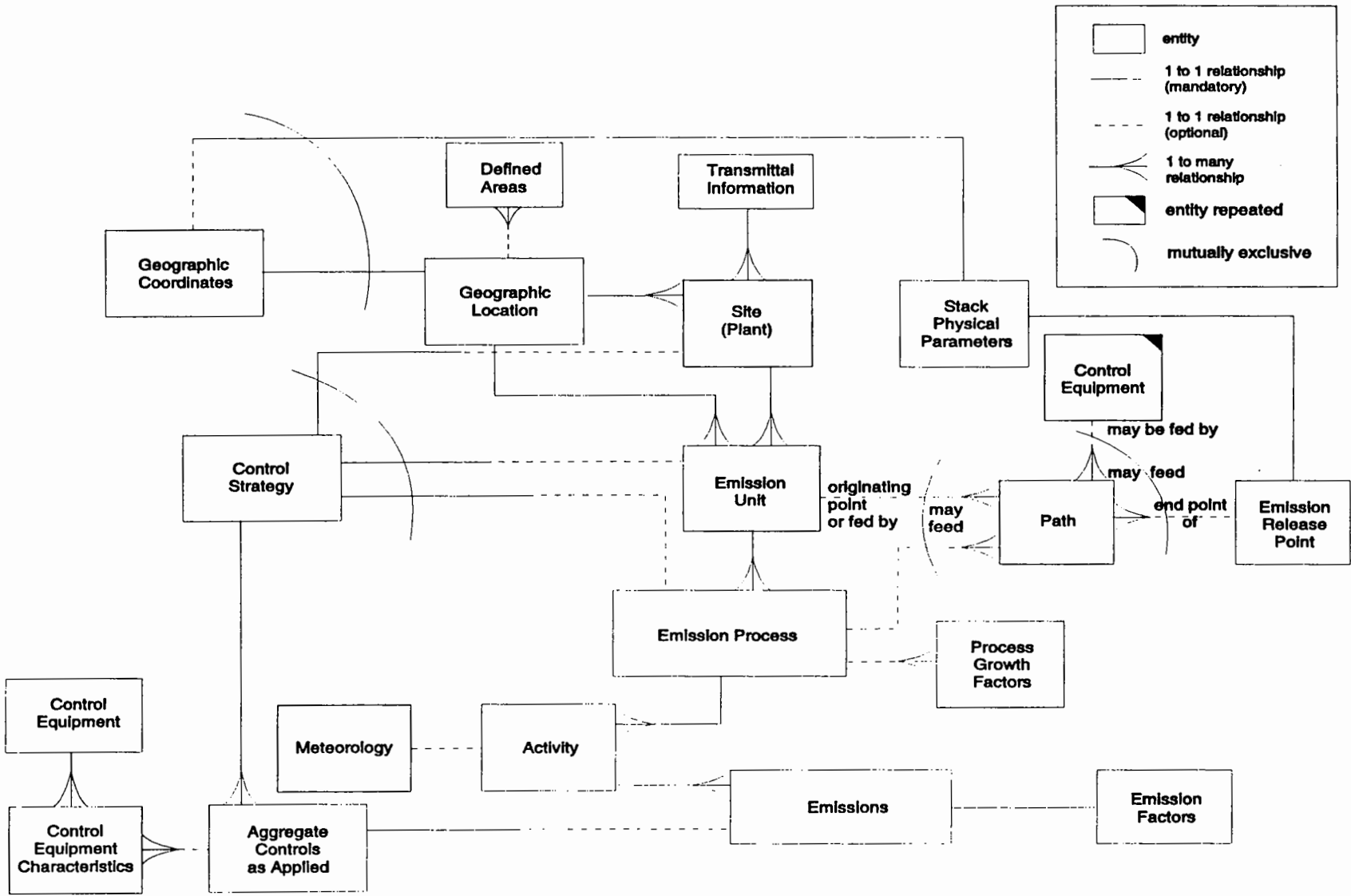


Figure 2. Point Source ERD



Figure 3. Area Source and Non-Road Mobile Source ERD

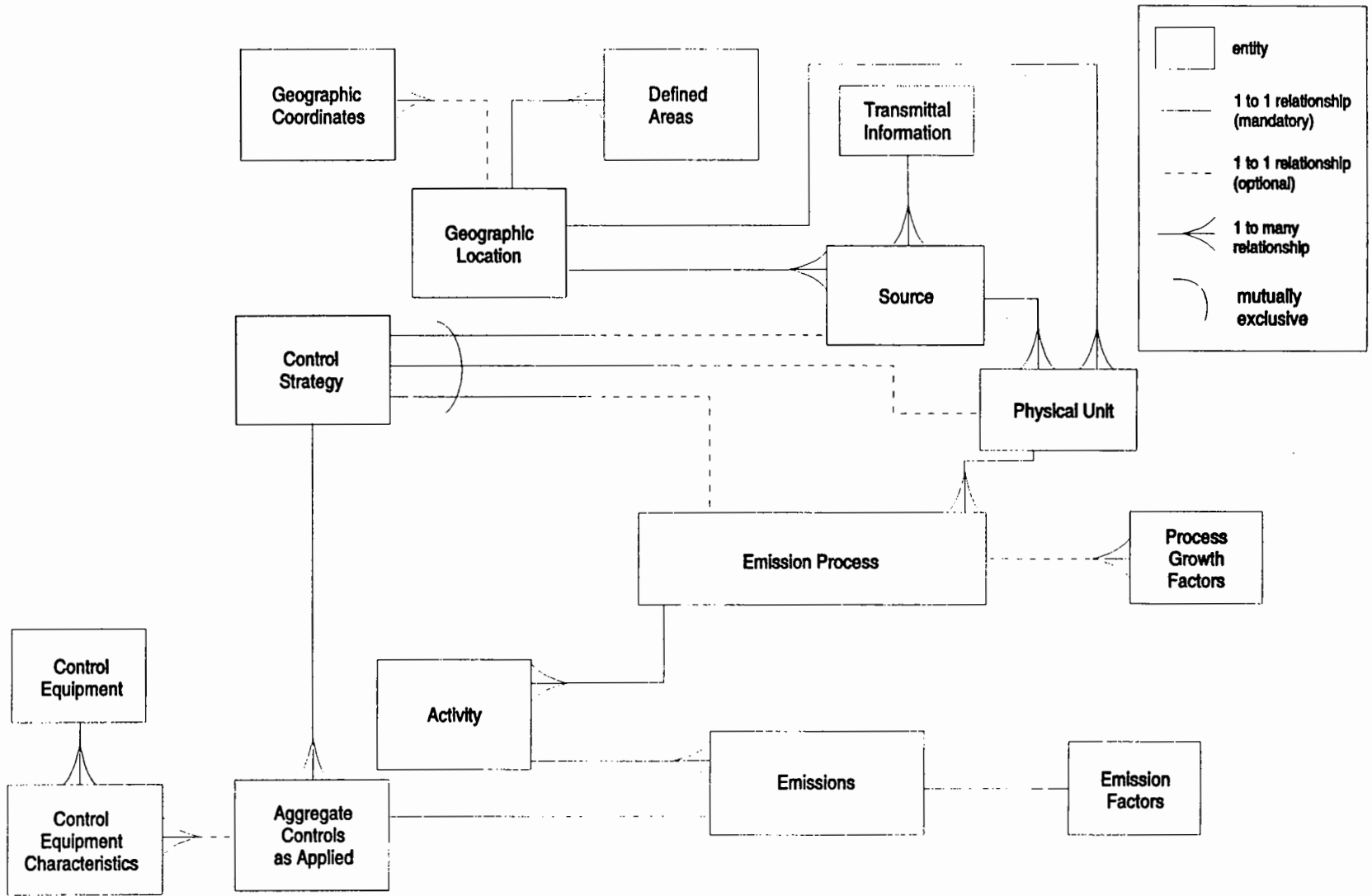


Figure 4. Mobile Source ERD

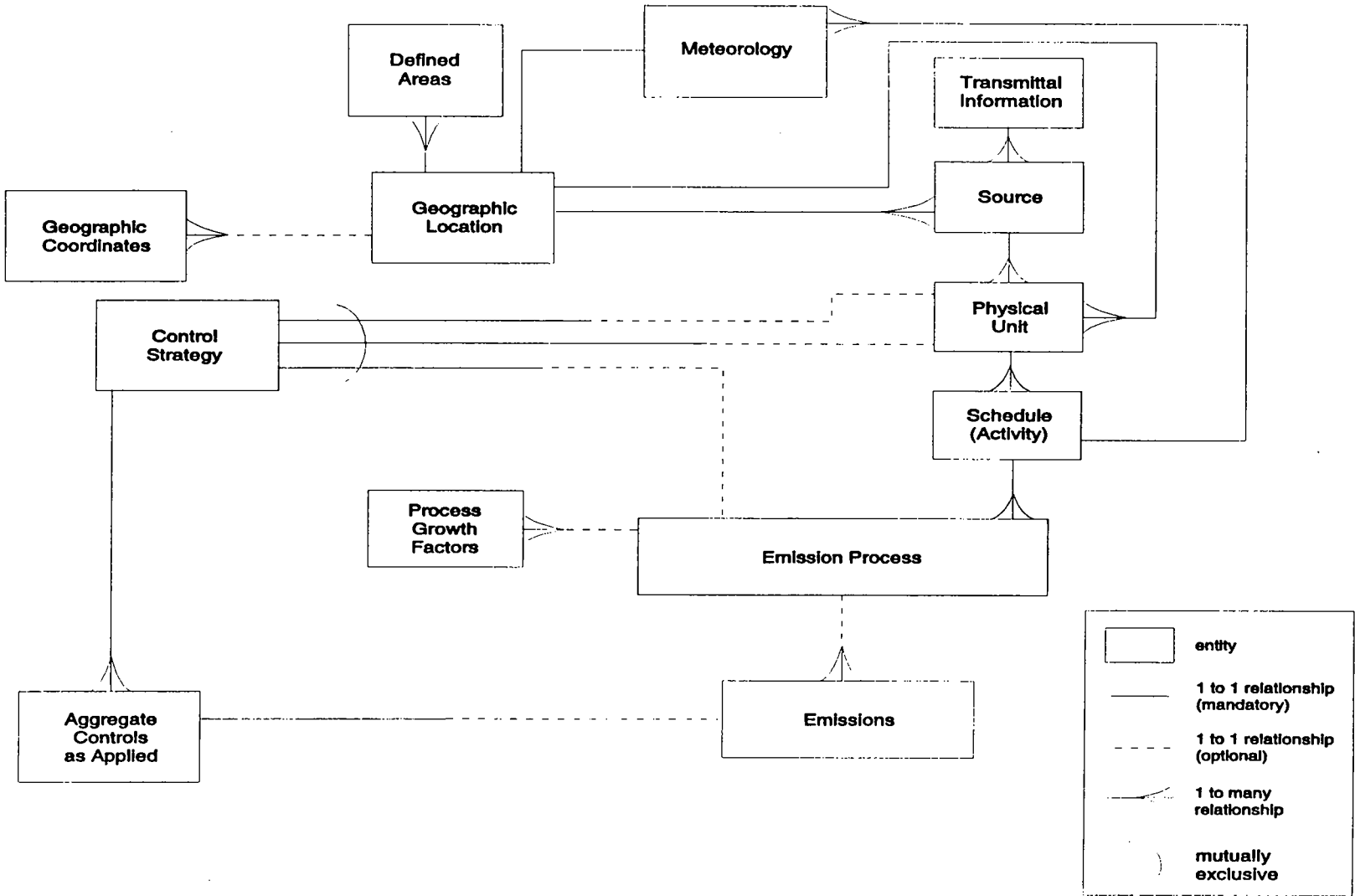
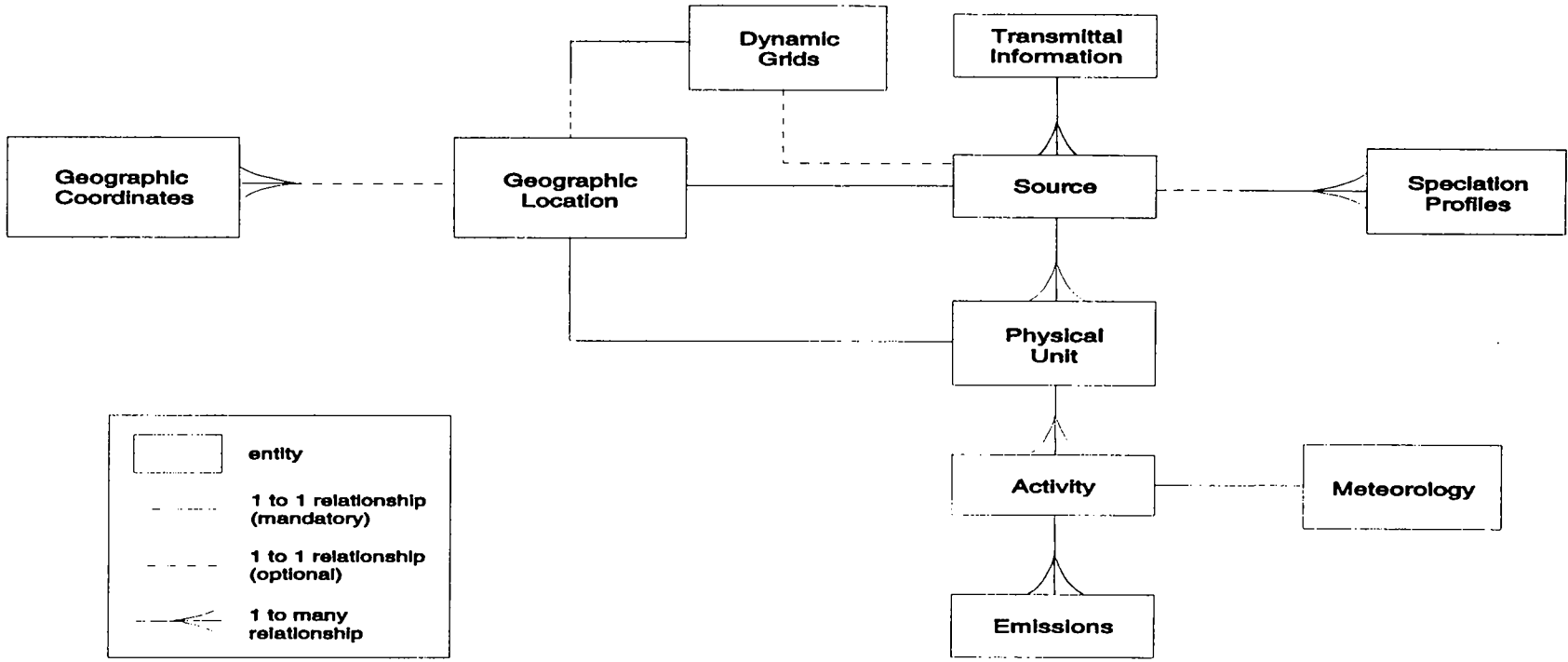


Figure 5. Biogenic Source ERD



**Figure 6. Site Attribute Table**

Data Attribute	Notes
Site UID	Site Unique Identifier Code
Site Geographic UID	Site Geographic Unique Identifier Code
Transaction ID	Data Transaction Identifier
Site Name	
Physical Street Address	
Physical City	
Physical State/Province	
Physical Country	
Physical Zip Code	
Mailing Street Address	
Mailing City	
Mailing State/Province	
Mailing Country	
Mailing Zip Code	
SIC	FTC Standard Industrial Classification Code
Number of Employees	
Description	
Federal ID Code	AFS Plant ID
Federal ID Code #2	EPA Key Identifier Code
ORIS Code	The Office of Regulatory Systems (ORIS) code (Department of Energy)
Jurisdiction	
State/Local Site ID	
Confidentiality Indicator	
Jurisdiction (Secondary)	
Dun & Bradstreet Number	

**Figure 7. Emission Unit Attribute Table**

<b>Data Attribute</b>	<b>Notes</b>
Unit UID	
Site UID	
Physical Unit Geographic UID	
Federal ID Code	AFS Point ID
Description	
Emission Unit Type Code	
Number of Units	
Confidentiality Indicator	
Date Installed/Modified	
Status	
Rule Applicability	
Design Capacity	
Design Capacity Units	
State/Local Emission Unit ID	
Status	

**Figure 8. Emission Process Attribute Table**

Data Attribute	Notes
Process UID	
Site UID	
Unit UID	
Start Date/Time	
End Date/Time	
Description	
SCC	
Process SIC	
Federal ID Code	AFS Segment ID
AMS Code	
Winter Throughput Percentage	
Spring Throughput Percentage	
Summer Throughput Percentage	
Fall Throughput Percentage	
Annual Average Hours Per Day	
Annual Average Days Per Week	
Annual Average Hours per Year	
Annual Average Weeks Per Year	
Heat Content	
Sulfur Content	
Ash Content	
Confidentiality Indicator	
Material	
Material Description	
State/Local Process ID	

**Figure 9.** Activity Attribute Table


<b>Data Attribute</b>	<b>Notes</b>
Site UID	
Unit UID	
Process UID	
Start Date/Time	
Process Rate/Throughput	
Confidentiality Indicator	
Maximum Actual Throughput	
End Date/Time	
Throughput Method Code	
Reliability Indicator	Data Attribute Rating System (DARS)score
Unit of Measure Text	
Unit of Measure Expression	
Period Average Hours Per Day	
Period Average Days Per Week	
Period Average Hours Per Year	
Period Average Weeks Per Year	

**Figure 10. Emissions Attribute Table**

Data Attribute	Notes
Unit UID	
Process UID	
Start Date/Time	
End Date/Time	
Pollutant Code	Such as VOC, PM, NO <sub>x</sub>
Emission Type	Such as Actual, Potential, Maximum
Numeric Value	
Unit of Measure Text	
Unit of Measure Expression	
Confidentiality Indicator	
Control Status	Controlled or Uncontrolled
Emission Estimation Method Code	Such as Stack Test, Emission Factor
Reliability Indicator	DARS score



TECHNICAL REPORT DATA

1. REPORT NO. <b>EPA/600/A-97/101</b>	2.	3. REC  PB98-116288
4. TITLE AND SUBTITLE Development of the EIIP Emission Inventory Data Model for purposes of facilitating data transfer between the states and the U.S. EPA	5. REPORT DATE	
	6. PERFORMING ORGANIZATION CODE	
7. AUTHOR(S) John Slade <sup>1</sup> , William Benjey <sup>2</sup> , Andrew Blackard <sup>3</sup> , and Michael Pring <sup>3</sup>	8. PERFORMING ORGANIZATION REPORT NO.	
9. PERFORMING ORGANIZATION NAME AND ADDRESS <sup>1</sup> Pennsylvania Department of Environmental Resources P.O. Box 8468, 400 Market St., Harrisburg, PA 17105 <sup>2</sup> Atmospheric Sciences Modeling Division, ARL, NOAA, Research Triangle Park, NC (on assignment to National Exposure Research Laboratory, US EPA, Research Triangle Park, NC) <sup>3</sup> Eastern Research Group, Inc., 1600 Perimeter Park, Suite 300, Morrisville, NC 27560	10. PROGRAM ELEMENT NO.	
	11. CONTRACT/GRANT NO.	
12. SPONSORING AGENCY NAME AND ADDRESS NATIONAL EXPOSURE RESEARCH LABORATORY OFFICE OF RESEARCH AND DEVELOPMENT U.S. ENVIRONMENTAL PROTECTION AGENCY RESEARCH TRIANGLE PARK, NC 27711	13. TYPE OF REPORT AND PERIOD COVERED	
	14. SPONSORING AGENCY CODE	
15. SUPPLEMENTARY NOTES		
16. ABSTRACT <p>The Emission Inventory Improvement Program (EIIP) was formed with the cooperation of the State and Territorial Air Pollution Program Administrators/Association of Local Air Pollution Control Officials (STAPPA/ALAPCO) as a cooperative effort between State and local agencies, industry, and EPA to improve emission estimates and establish a National Standard for Emission Data Exchange. Emission data are needed for a variety of purposes, including regional air quality modeling, air quality trend planning, trends analysis, and for informing the public. The Data Management Committee (DMC) of the EIIP was charged with the goal of developing the data exchange standard, and undertook the task of developing a data model on which to base the data exchange format. The purpose of developing the data model was to establish a standardized set of data relationships which reflect physical reality and ensure accurate and consistent reporting of emissions and emission related data.</p>		
17. KEY WORDS AND DOCUMENT ANALYSIS		
a. DESCRIPTORS	b. IDENTIFIERS/ OPEN ENDED TERMS	c. COSATI
18. DISTRIBUTION STATEMENT	19. SECURITY CLASS (This Report)	21. NO. OF PAGES
	20. SECURITY CLASS (This Page)	22. PRICE