

Acknowledgements

The Core Reference Model (CRM) Workgroup is comprised of participants from EPA and States, along with contractor support. Core Reference Model Workgroup members included:

State, ECOS, and US EPA Members	Organization
Tom Aten	Wisconsin DNR
Michael Beaulac (Project Leader)	Michigan DEQ
Mary Blakeslee	ECOS
Dennis Burling	Nebraska DEQ
Tim Crawford	US EPA / DSB
Pat Garvey	US EPA / OIC
Sarah Hisel-McCoy	US EPA / OEI
Gail Jackson	Pennsylvania DEP
David Kempson	Arizona DEQ
Tom Lamberson	Nebraska DEQ
Dennis Murphy	Delaware DNR
Sandy Smith	Missouri DNR
Linda Spencer	US EPA – DSB

Contractors	Organization
Sarah Calvillo	Ross & Associates
Greg Carey	enfoTech & Consulting Inc.
Tony Jeng	enfoTech & Consulting Inc.
Louis Sweeny	Ross & Associates
Douglas Timms	enfoTech & Consulting Inc.

Table of Contents

1	Introduction.....	5
2	Background and Approach.....	7
2.1	What is the Core Reference Model?	7
2.2	Role of the CRM in the Exchange Network	8
2.3	CRM Development Timeline.....	10
3	Overview of the Core Reference Model	11
3.1	The Concept	11
3.2	Data Element.....	12
3.3	Data Block	12
3.4	Major Data Group	13
4	Core Reference Model Inventory	17
4.1	Inventory Overview	17
4.2	Data Block Inventory Details.....	19
4.3	Data Blocks Removed.....	26
4.4	Major Data Group Inventory Details	28
5	CRM Inventory Analyses.....	48
5.1	Reusable Data Block Analysis.....	48
5.2	CRM / XML Schema Comparison Analysis	49
5.3	Sample Uses of the CRM to Data Standard Development	50
6	Example Applications of the Core Reference Model.....	51
6.1	Facility-to-State Data Flow (Environmental Reports and Forms).....	51
6.2	State-to-USEPA Data Flow	52
6.3	Facility-to-USEPA Example: RCRA Permit Application	52
6.4	Other Uses.....	53
7	Recommended Future Steps	55
7.1	Resolve Inconsistencies Between CRM and EDSC Data Standards	56
7.2	Define CRM's role in the Exchange Network.....	56
7.3	Launch CRM Phase III Development (based on item above)	57
7.4	Responsible Parties for CRM Maintenance.....	58
8	Appendices.....	59
8.1	Definitions and Abbreviations	59
8.2	References.....	60

Listing of Diagrams and Tables

Diagram 1: Relationship of Three Major Conceptual Components for the CRM.....	7
Diagram 2: Role of CRM and Shared Schema Components in XML Development.....	9
Diagram 3: CRM Development Timeline	10
Diagram 4: CRM Legend.....	11
Diagram 5: Example Data Block.....	13
Diagram 6: Example Major Data Group	14
Table 1: Major Data Group Inventory Detail.....	15
Diagram 7: Major Data Group Inventory (The Big Picture)	18
Table 2: Complete Data Blocks Inventory Detail	19
Table 3: Data Blocks Removed from CRM Version 2.0	26
Diagram 8: Major Data Group of Compliance Result (CR)	29
Diagram 9: Major Data Group of Contact (C)	30
Diagram 10: Major Data Group of Enforcement (E).....	31
Diagram 11: Major Data Group of Environmental Accident Event (EAE)	32
Diagram 12: Major Data Group of Environmental Notice (EN)	33
Diagram 13: Major Data Group of Facility (F)	34
Diagram 14: Major Data Group of Grant (G).....	35
Diagram 15: Major Data Group of License (L).....	36
Diagram 16: Major Data Group of Monitoring Ambient (MA)	37
Diagram 17: Major Data Group of Monitoring Compliance (MC).....	38
Diagram 18: Major Data Group of Monitoring Emergency (ME)	39
Diagram 19: Major Data Group of Permit (P)	40
Diagram 20: Major Data Group of Permit (P) - Continued	41
Diagram 21: Major Data Group of Release (R).....	42
Diagram 22: Major Data Group of Reference Method and Factor (RMF)	43
Diagram 23: Major Data Group of Reporting (RPT)	44
Diagram 24: Major Data Group of Substance (S).....	45
Diagram 25: Major Data Group of Spatial Data (SD)	46
Diagram 26: Major Data Group of Source (SR).....	47
Diagram 27: Facility Schema Represented by Data Blocks.....	49

1 Introduction

The Core Reference Model (CRM) is a high-level depiction of major groupings of environmental data and their relationships. It was created to provide federal, state, and tribal environmental agencies with guidance for consistently building and sharing environmental data on the Exchange Network. By providing a high-level environmental data model that accommodates a variety of environmental topics, the CRM facilitates the creation of Data Exchange Templates (DET) such as XML schema for any variety of environmental data exchanges that share common components. By providing a complete model of environmental information, the CRM also provides the Environmental Data Standards Council (EDSC) with opportunities to identify new data standards as well as guidance on the structuring of data standards.

In addition to providing a high-level data model, the CRM, in conjunction with data standards developed by the EDSC, facilitates the creation of Shared XML Schema Components (SSC), which are basic XML building blocks that can be used by those designing, revising, or expanding environmental information exchanges via XML schema creation.

The key objectives for the CRM include:

- Describing a high-level overview of environmental data, organized into a meaningful model that promotes the creation of consistent and related Data Exchange Templates (DET)
- Providing basic building blocks for Partners to use in data exchange projects promoting interoperability among data flows
- Discouraging the creation of redundant or conflicting XML schema development efforts
- Identifying areas for potential data standardization
- Identifying certain key Data Elements required for each data schema to promote DET harmonization
- Creating a tool for Exchange Network managers and members to carry-out their respective roles to guide/manage and assist future XML schema development

This document provides an overview of the *Core Reference Model (CRM)*, highlighting its purpose and value in supporting the development of an integrated data exchange infrastructure among state and federal environmental agencies. Details of the CRM are provided, including the structure and meaning of the current model.

The document also provides a background into how the CRM was developed and its relationship to other Exchange Network tools, notably the Environmental Data Standards Council (EDSC) data standards and Shared Schema Components (SSC). Finally, recommendations are provided that will allow the CRM to continue to be a relevant tool for environmental data exchange development.

Two companion documents have been created along with the Core Reference Model II:

- **SSC Usage Guide:** introduces the Exchange Network Shared Schema Components (SSC), illustrates the benefits of using sharable schema components based on approved EDSC data standards as an alternative to XML schema developed without such standards, and provides detailed guidance to XML schema developers on how they can incorporate the SSC into their data flow XML schema.

- **SSC Technical Reference:** provides a detailed technical representation of the Shared Schema Components (SSC). For each SSC, the elements that are referenced and their details (namespace, type, attributes, facet restrictions, and annotations) are provided.

2 Background and Approach

2.1 What is the Core Reference Model?

The CRM Workgroup has sought to create the common business framework for sharing environmental information on the Exchange Network. This business framework is represented by three distinct conceptual components as follows:

- **Data Element:** A single unit of data that cannot be divided and still have useful meaning. Data Elements in the CRM may directly correspond to those found in existing data standards, XML schema, database field names, and entities found in the Environmental Data Registry (EDR).
- **Data Block:** A grouping of related Data Elements and other Data Blocks¹ that can be used and reused among different information flows. An example Data Block is Agency Identification, which includes the component Data Elements such as Agency Identifier, Agency Name, Agency Type, and Facility Management Type.
- **Major Data Group:** a logical grouping of related Data Blocks that fully describe business areas, functions, and entities where EPA and its Partners have an environmental interest. Major Data Groups provide a logical path for locating and retrieving Data Blocks. An example Major Data Group is Contact, which may include Data Blocks such as Individual Identity and Mailing Address.

These ideas are illustrated in the diagram below:

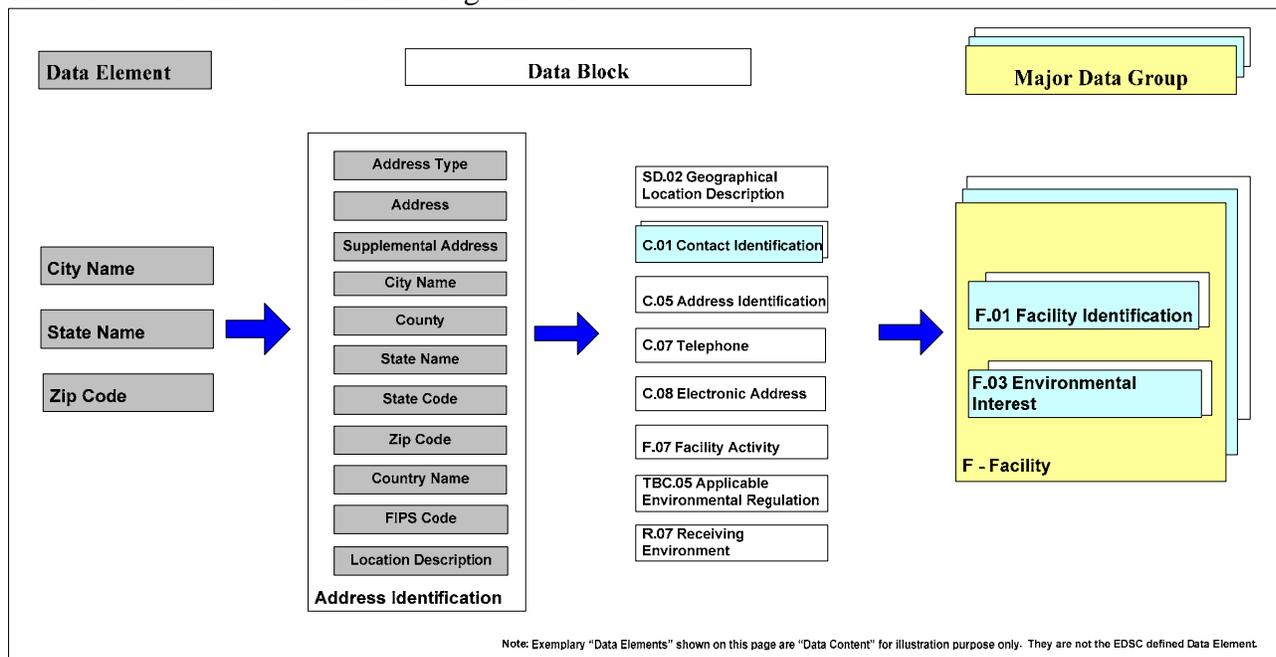


Diagram 1: Relationship of Three Major Conceptual Components for the CRM

¹ The Core Reference Model version 1.0 distinguished Data Blocks (consisting solely of data elements) from Compound Data Blocks (consisting of data elements and other smaller data blocks). In CRM Version 2, the concept of Compound Data Blocks has been removed and so the Compound Data Blocks are now identified as Data Blocks.

The current data standards adopted by the EDSC are groupings of Data Elements. This is similar to the use of Data Blocks used in CRM. However, some existing data standards may not match the CRM Data Block approach and may need to be restructured or harmonized once a set of CRM Data Blocks have been agreed upon.

2.2 Role of the CRM in the Exchange Network

Environmental agencies are working on numerous data exchange efforts, including both internal and external exchanges with other agencies. The Exchange Network was conceptualized and developed to enhance the way in which information is stored and shared among tribal, state, and federal environmental agencies. The Exchange Network is the culmination of several directed efforts and the primary focus for the recent US EPA Network Grant awards to the States, Tribes, and Territories. Partners commit to change the way data is exchanged and to build their individual capacities to make essential data accessible.

Early in the development of the Exchange Network, the Core Reference Model was identified by a variety of Exchange Network oversight bodies as a key component essential to the promotion of consistent data exchanges. Because the vision of the Exchange Network is one in which data shared on the Network is easily understood by all Partners, there is a primary goal is to achieve interoperability among all Partners via a common business framework that facilitates the sharing of data. This common business framework is achieved through cooperation between three key Exchange Network components: environmental data standards, XML schema design guidelines, and the CRM.

1. EDSC Environmental Data Standards:

Standards are a fundamental cornerstone of e-Government, the Exchange Network, and systems integration. Data standards must be in place to enable efficient and integrated flow of data across the Exchange Network. The EDSC was created by the IMWG in 2000 to promote the efficient sharing of environmental information among the Partners and other parties through the development of data standards. The EDSC's objective is to foster the development of data standards that support the Exchange Network.

2. XML design guidance and rules:

The Exchange Network provides XML design guidance through a variety of means. The *XML Design Rules and Conventions* document provides technical recommendations to the Partners on XML schema development. This document also provides techniques for extending the core XML schema modules to meet special requirements of future users. Additional XML guidance such as XML Namespace guidance is also available to ensure consistent XML schema development.

3. Create a mechanism to facilitate construction and reuse of Exchange Network Shared Schema Components:

The CRM defines key Data Blocks as a collection of commonly used data elements. Reusable XML schema modules (called Shared Schema Components (SSC)) are also developed as a direct representation of the CRM and data standards that can be used by Partners in the development of XML schema.

These three efforts provide the common language for sharing data on the Exchange Network. The data standards provide the vocabulary, XML Design Rules and Conventions are the

grammar and syntax rules, and the CRM defines the topics that the Partners will discuss. The contributions from each effort are illustrated on the diagram shown on the next page:

Role of CRM and Shared Schema Components in XML Development

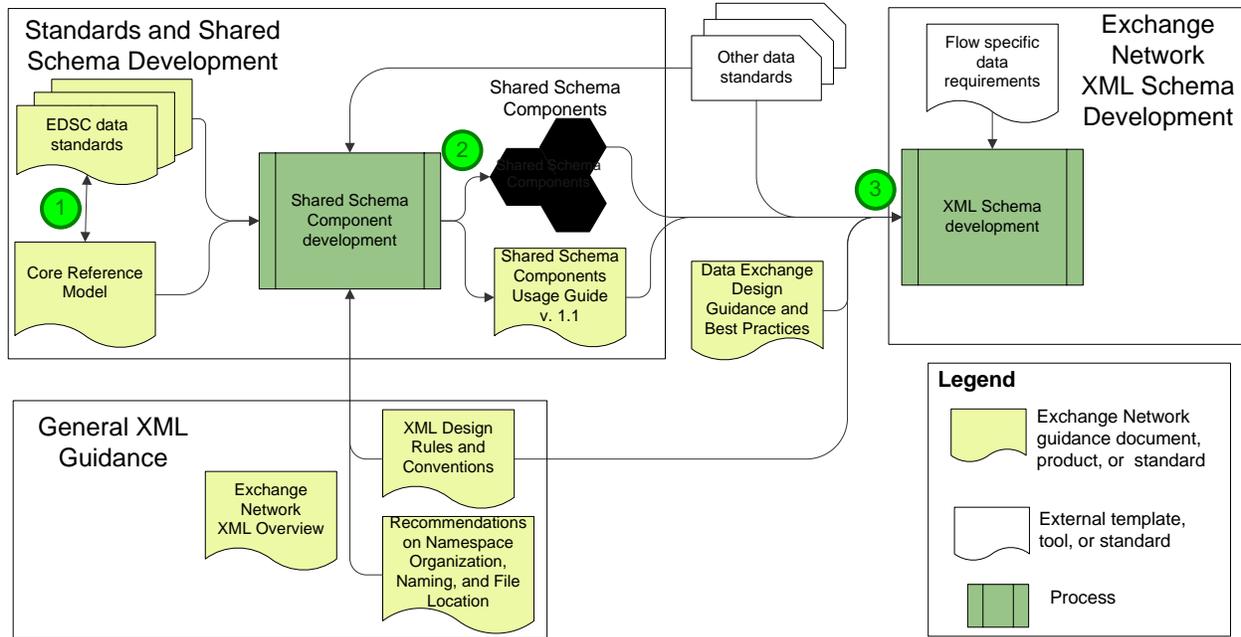


Diagram 2: Role of CRM and Shared Schema Components in XML Development

Three key aspects of this diagram are described below:

- 1 EDSC Data Standards / CRM interaction:** Data standards are developed by the EDSC based partially on guidance from data modeling concepts defined in the Core Reference Model. The Core Reference Model is in turn influenced and refined based on data standards development from the EDSC.
- 2 Shared Schema Components Development:** When EDSC data standards are finalized, shared XML schema components (SSC) are created that provide reusable XML schema that organize related data elements common to multiple environmental data flows. They incorporate Environmental Data Standards Council (EDSC) data standards for data element grouping, data element names, and definitions
- 3 XML Schema Development:** As shown in the diagram, Exchange Network XML schema are created based on Shared Schema Components (SSC), general XML guidance, external data standards, and flow-specific requirements. Because SSCs are created from CRM, CRM ultimately play a role in the development of Exchange Network XML schema.

2.3 CRM Development Timeline

The CRM has been developed in two Phases over the last three years by the Core Reference Model Workgroup, as a part of the Exchange Network. The following diagram depicts the development timeline of the Core Reference Model and related tools.

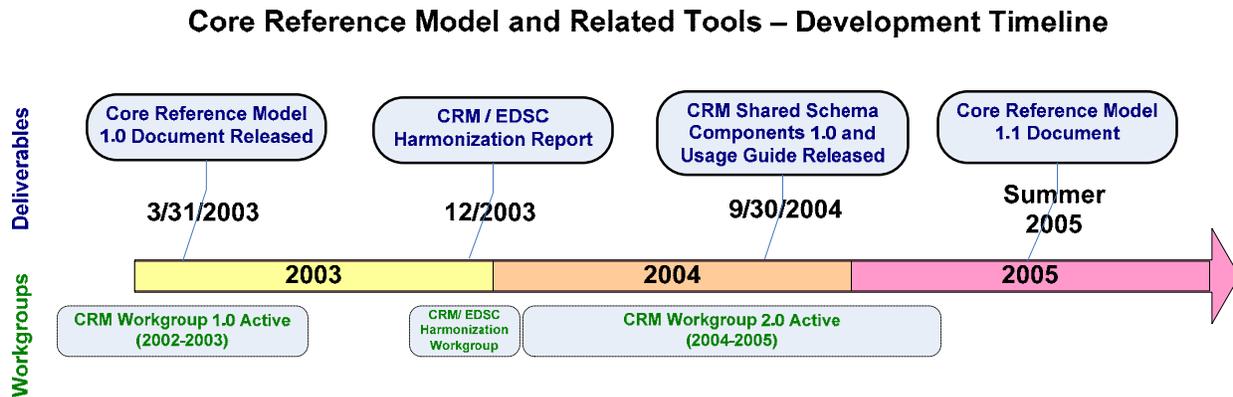


Diagram 3: CRM Development Timeline

Core Reference Model Phase I Workgroup:

The primary objective of the Phase I workgroup was to create and articulate the Core Reference Model. This was accomplished via the publication of the *Core Reference Model for the Environmental Information Exchange Network* document, version 1.0, in March 2003. This document introduced the concept of a modular environmental data model by providing a high-level depiction of the major groupings of environmental data and their relationships.

The CRM Workgroup met with members of the Environmental Data Standards Council (EDSC) in October 2003 to harmonize data element names, blocks/groups and definitions between the two entities resulting in a revised version of the CRM. In addition to the high-level depiction, the CRM document also introduced the idea of creating reusable XML schema for Exchange Network use, which led to the activities conducted by the Phase II workgroup in 2004.

Core Reference Model Phase II Workgroup:

The Core Reference Model Phase II Workgroup convened in April 2004 with the goal of creating shared XML schema using as the basis the data blocks and major data groups identified during Phase I and subsequent harmonization efforts with EDSC.

The Phase II Workgroup also updated the Core Reference Model for the Environmental Information Exchange Network document to version 2.0 in July 2005. The document was updated to reflect a more refined understanding of the Core Reference Model that had emerged since the publication of version 1.0 back in March 2003. Changes include an update to reflect recent changes to the data standards, closer ties between data standards and the Core Reference Model, updated references, and a more streamlined representation of the model.

3 Overview of the Core Reference Model

3.1 The Concept

The CRM is a high-level depiction of major groupings of environmental data and their relationships. These components include Data Elements, Data Blocks, and Major Data Groups. Throughout the development of the CRM, two approaches have been used: the “coarse-to-fine grain” approach to group environmental data from a top-down business process perspective, and the “fine-to-coarse grain” approach used to examine the data elements and how they are used as a method to validate the grouping of the elements into the appropriate Data Blocks and Major Data Groups.

The following diagram demonstrates the top-down, or “coarse-to-fine” view of the CRM. At the top of the diagram, the coarsest view is the Major Data Group. Conversely, at the bottom of the diagram, the finest view is the Data Element. While the CRM Workgroup acknowledges that other components exist that are of finer detail than the Data Element level (such as valid lists of values for code lists), the Workgroup decided to not focus beyond the Data Element level at this time. Data Elements included in the CRM are for illustration purposes and are not intended to be a complete list.

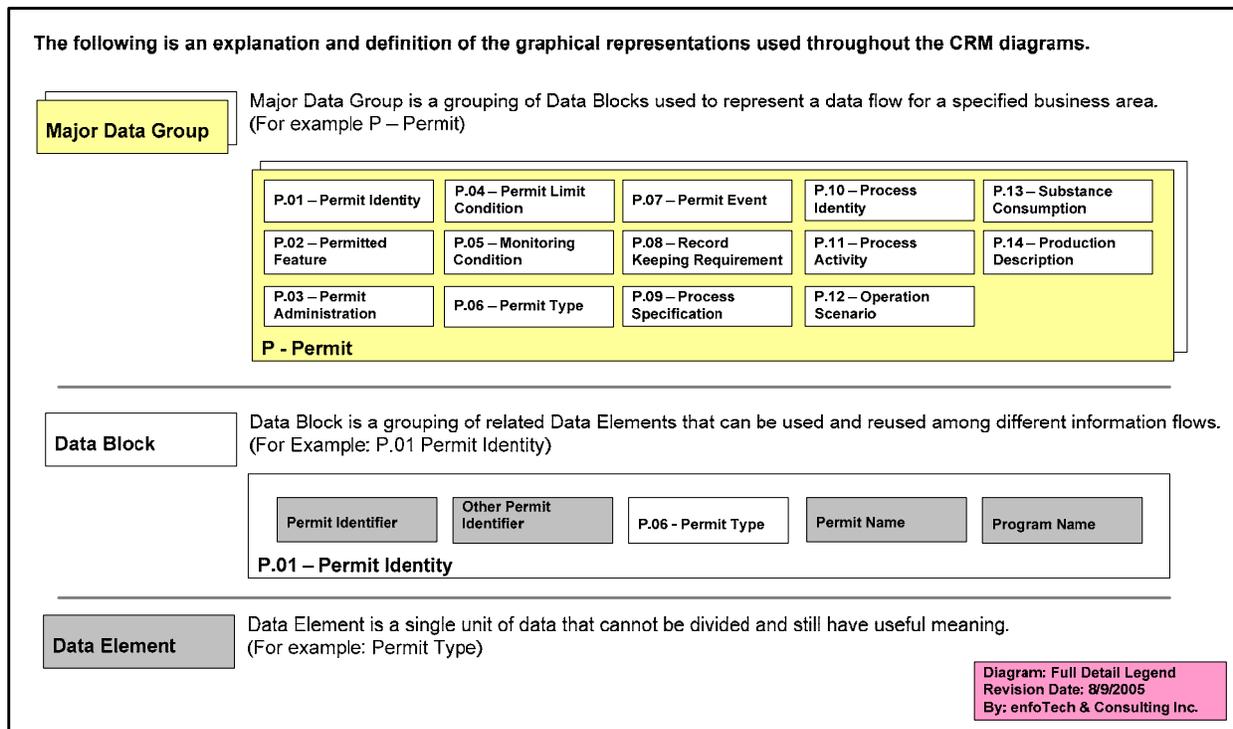


Diagram 4: CRM Legend

The following sections demonstrate the “fine-to-coarse” view of the CRM.

3.2 Data Element

A Data Element is the most basic unit of data exchange. A Data Element is a single unit of data that cannot be divided and still have useful meaning. Data Elements in the CRM may directly correspond to those found in existing data standards, XML schema, database field names, and entities found in the Environmental Data Registry (EDR). Example Data Elements are individual components of a Mailing Address, such as Mailing Address City Name and Address Postal Code.

In the initial phase of the CRM Project, the Workgroup made an attempt to list “example” Data Elements for each Data Block for the purpose of communicating Data Block definitions. These elements are captured in the CRM Version 1 document. Since the CRM Version 1 document was released, additional elements had been identified as part of a harmonization effort with the EDSC. As a result, additional data elements have been supplied. Although this document identifies data elements, the Workgroup does not intend to identify all Data Elements for this phase of the Project. Instead, representative Data Elements are used in the CRM document and diagrams which are used to help illustrate and reinforcement the meaning of the Data Blocks to which the Data Elements are associated. They are not necessarily the complete listing of data elements approved by the EDSC for a particular data block.

3.3 Data Block

A Data Block is a grouping of related Data Elements and other data blocks² that can be used and reused among different information flows. A Data Block must contain more than one child element. An example Data Block is *Mailing Address*, which consists of the component Data Elements Mailing Address Text, Supplemental Address Text, Mailing Address City Name, and Address Postal Code, and the smaller Data Blocks State Identity and Country Identity. Therefore, *Mailing Address* is an example Data Block that is constructed of both Data Elements and smaller Data Blocks.

Currently, the CRM adopts the following naming convention for Data Blocks:

Identifying Letter from Major Data Group.00 Data Block Name

For example, the Data Block, C.04 Mailing Address, is composed of:

- Identifying Letter from Major Data Group = C (i.e. Contact)
- Two Digit Unique Sequence Number within its Major Data Group³ = 04
- Data Block Name = Mailing Address

² The Core Reference Model version 1.0 distinguished Data Blocks (consisting solely of data elements) from Compound Data Blocks (consisting of data elements and other smaller data blocks). In CRM Version 2, the concept of Compound Data Blocks has been removed and so the Compound Data Blocks are now identified as Data Blocks.

³ Although there is no rule regarding the ordering of data blocks within a particular Major Data Group, blocks are typically ordered in a manner that is logical with respect to business logic. For example, in the *Contact* Major Data Group, the Data Block *C.01: Individual Identity* precedes *C.04: Mailing Address* since an individual is usually described before the individual’s mailing address. This approach is used for all Data Block ordering.

The following diagram depicts an example Data Block, comprised of both data elements and smaller Data Blocks:

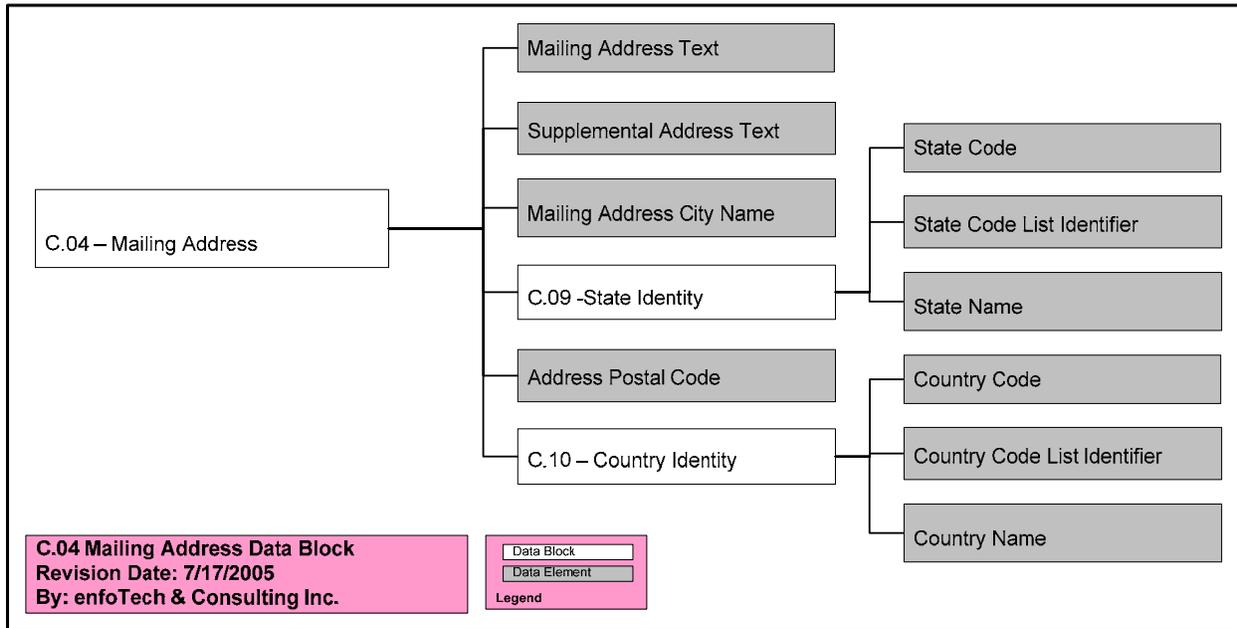


Diagram 5: Example Data Block

A complete inventory of the Data Blocks defined in the CRM, organized by Major Data Group, is provided in Section 4.

3.4 Major Data Group

The CRM groups related Data Blocks into Major Data Groups to more fully describe high level business areas, functions, and entities where EPA and its Partners have an environmental interest. Users of the CRM may find these helpful in designing flow schema and information systems. The assembly of Data Blocks into Major Data Groups is intended to serve as a guide to reflect common relationships among data. However, these are not fixed relationships – users should rely primarily on Data Blocks to build their databases and flow schema.

Below is the graphical representation for an example Major Data Group: Contact (C). This Major Data Group consists of eleven Data Blocks.

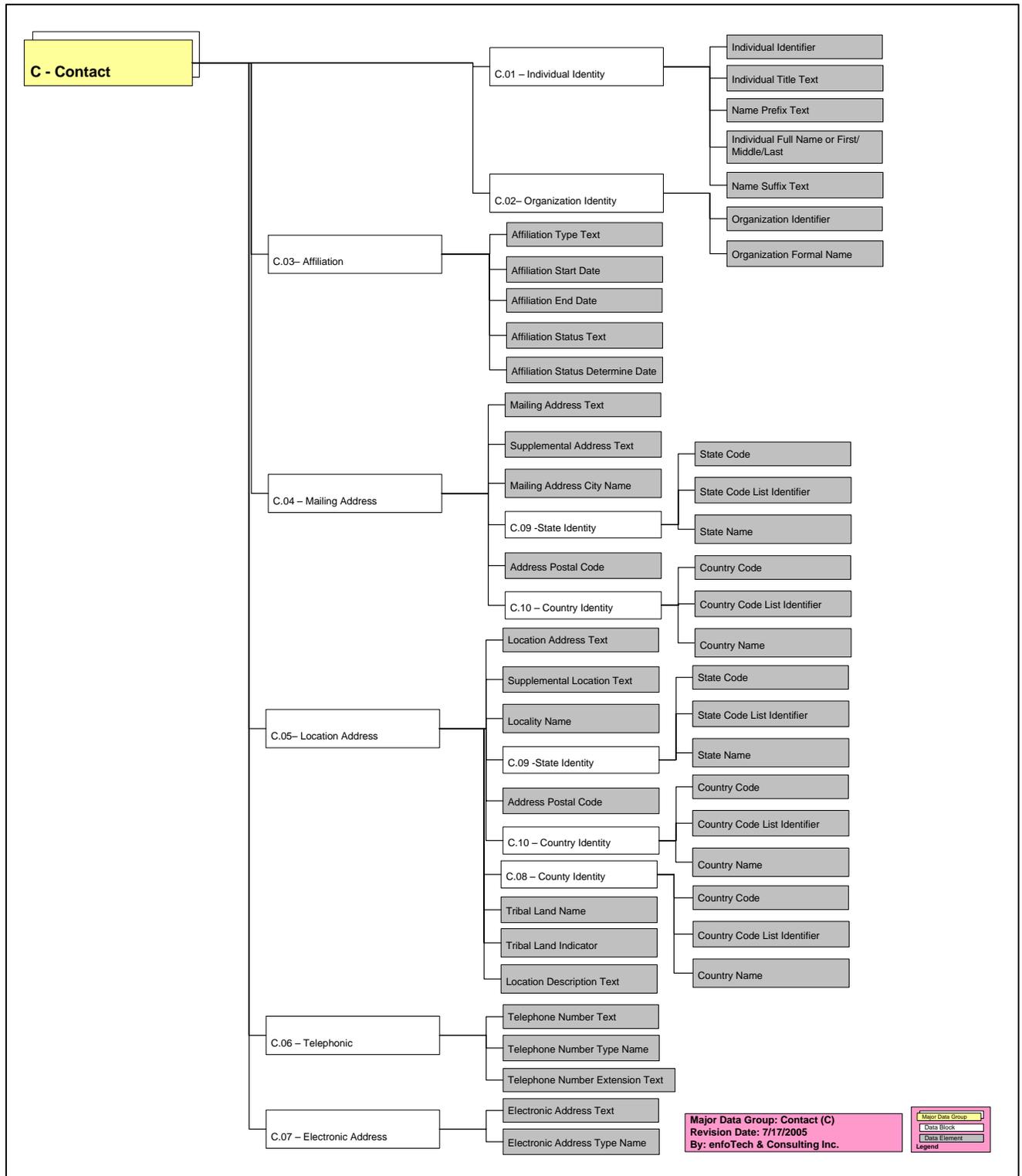


Diagram 6: Example Major Data Group

The following Major Data Groups are defined in the CRM:

Table 1: Major Data Group Inventory Detail

Major Data Group ID	Major Data Group Name	Major Data Group Description	Created During CRM Phase
C	Contact	Basic information needed for contacting an individual or organization. This might include individual name, organization name, address, telephone, etc.	I
CR	Compliance Result	Information about the regulatory assessment of a given environmental activity. A compliance result could be generated from a given environmental release when comparing against a permit; generated by an inspection or monitoring conducted by agencies; a result of “non responsive” from the permittee (e.g., failure to submit a DMR report or late submission).	I
E	Enforcement	Actions taken by a regulatory agency against the regulated facility as a result of non-compliance with environmental laws or permits. This might include compliance schedule, consent order, fine, penalty, etc.	I
EAE	Environmental Accident Event	An event occurred that was un-expected or un-permitted that could have a potentially adverse impact to the environment or human health.	I
EN	Environmental Notice	Environmental advisories or statements issued by an agency that will provide advice/guidance to the general public on the use of resources (such as swimming restrictions, fish consumption, drinking water quality, etc.).	I
ESAR ⁴	Environmental Sampling Analysis & Results	Information related to environmental sampling, analysis, and results, covering both ambient and point source monitoring for all environmental media. This major data group also includes the characterization of monitoring locations.	II
F	Facility	An entity with name, location, and responsible party that may or may not be regulated.	I
G	Grant	Grant information such as payment order, award, proposal, deliverables, and area of application, among others. Scope of administration is the high-level data that the States and EPA exchange.	I
L	License	A regulatory document that contains activities for which the applicant (facility or individual) is allowed to perform. This may include specific professional practices or activities licensed (e.g., well-drilling, asbestos removal, stack testing, etc.)	I

⁴ The ESAR Major Data Group was introduced into the CRM Version 2.0 to reflect the new data standard published by the EDSC; and was codified with the release of several Shared Schema Components in late 2004. However, this Major Data Group currently overlaps with significant portions of the previous MA (Monitoring Ambient), MC (Monitoring Compliance, and ME (Monitoring – Emergency Response) Major Data Groups created during CRM Version 1.0. A harmonization of the ESAR Major Data Group with the MA, MC, and ME Major Data Groups is required and is still pending. Until that harmonization occurs, these overlapping Groups will continue to co-exist.

Major Data Group ID	Major Data Group Name	Major Data Group Description	Created During CRM Phase
MA ⁴	Monitoring-Ambient	Monitoring activities performed for locations in open public areas such as river water quality, ambient air quality, etc.	I
MC ⁴	Monitoring-Compliance	Monitoring activities performed as required by a given permit, rule, or regulation. This type of monitoring is associated with a facility and is performed as part of permit compliance.	I
ME ⁴	Monitoring-Emergency Response	Monitoring activities as a result of environmental accident, spill, emergency response, etc.	I
P	Permit	A regulatory document that provides terms and conditions to allow a facility to perform a series of defined environmental activities (best management practices) and/or to discharge substances of a specified quantity with conditions, monitoring requirements, reporting requirements, etc.	I
R	Release	Detailed information about a given substance that is released into the environment. This includes substance ID, release media, release location, release quantity, etc.	I
RMF	Reference Method and Factor	Information on standard reference methods used for analyzing substances from samples and published factors used for estimating releases.	I
RPT	Reporting	Information related to environmental reporting for both regulatory and non-regulatory reporting needs, including the characterization of report forms, reporting requirements, and report submission details.	II
S	Substance	Information used to define the substance used, manufactured, or released into the environment. This could be chemical, biological, radiological, or other un-categorized material that may be included in analysis. A substance may or may not be regulated. A substance could be regulated under multiple environmental regulations.	I
SD	Spatial Data	Description of boundaries that could be determined by nature, political interest, administrative management, and other considerations.	I
SR	Source	Information pertaining to the origin or starting point of emissions or discharges.	I
TBC	To Be Categorized (not currently in use)	Other general Data Blocks that do not have specific relationships to any Major Data Group named in the CRM inventory.	I

Detailed information for each of the Major Data Groups identified above in terms of their composition of Data Elements and Data Blocks, and changes since CRM v1.0 are provided in Section 4.

4 Core Reference Model Inventory

4.1 Inventory Overview

The following diagram displays the “Big Picture” of the CRM, including all Major Data Groups and Data Blocks within those groups:

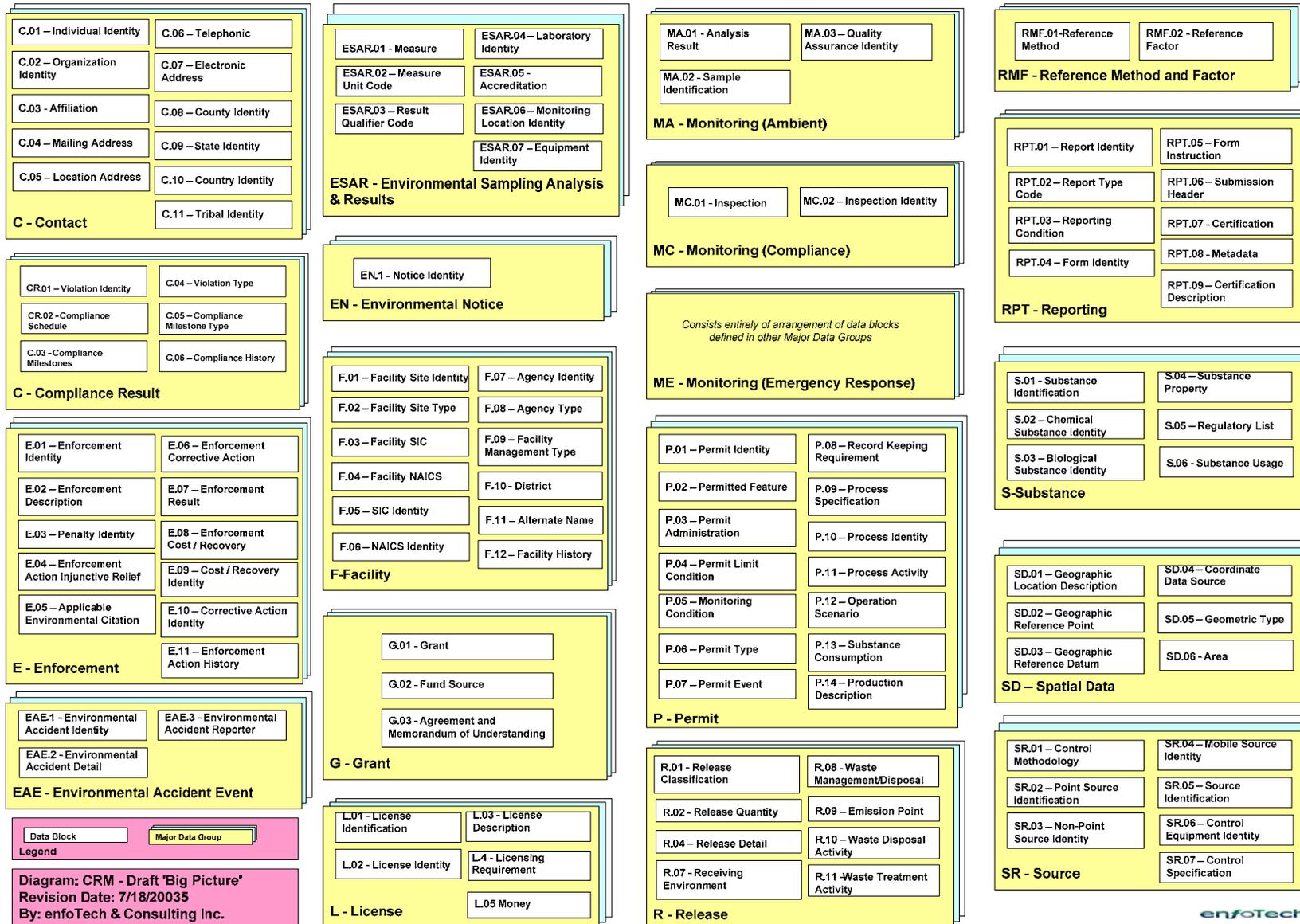


Diagram 7: Major Data Group Inventory (The Big Picture)

4.2 Data Block Inventory Details

The following table lists detailed information about the 116 Data Blocks identified/used in the CRM thus far:

Table 2: Complete Data Blocks Inventory Detail

Data Block ID	Data Block Name	Description	CRM I Data Block ID
C.01	Individual Identity	The particular terms regularly connected with a person so that you can recognize, refer to, or address him or her.	C.02
C.02	Organization Identity	The particular terms regularly connected with a unique framework of authority within which a person or persons act, or are designated to act, towards some purpose.	(new)
C.03	Affiliation ⁵	The relationship between an individual or organization and a facility, project, or actions.	C.04
C.04	Mailing Address	The standard address used to send mail to an individual or organization.	(new)
C.05	Location Address	The physical location of an individual or organization.	(new)
C.06	Telephone	An identification of a telephone connection.	C.07
C.07	Electronic Address	A location within a system of worldwide electronic communication where a computer user can access information or receive electronic mail.	C.08
C.08	County Identity	A code and associated metadata used to identify a U.S. county or county equivalent.	(new)
C.09	State Identity	A code and associated metadata used to identify a principal administrative subdivision of the United States, Canada, or Mexico.	(new)
C.10	Country Identity	A code and associated metadata used to identify a primary geopolitical unit of the world.	(new)
C.11	Tribal Identity	Basic information used for tribal entity identification (federally recognized tribes) and is not intended for the identification of geographic, demographic, or economic tribal areas. Examples may include tribal name, tribal code, etc.	C.03
CR.01	Violation Identity	Basic identification information used for defining a noncompliance with one or more legally enforceable obligations by a regulated entity, as determined by a responsible authority.	CR.03
CR.02	Compliance Schedule	Information about the types of activities leading to or resulting in a determination of a regulated facility returning to compliance.	CR.02
CR.03	Compliance Milestones	Information about the status of implementation, by Defendant/ Respondent, of compliance actions required as milestones included in a Final Order or other enforcement action resolution, including Injunctive Relief, Supplemental Environmental Projects (SEP), and Penalty or Cost Recovery payments required.	(new)
CR.04	Violation Type	A designator and associated metadata used to identify a type or class of violation.	(new)
CR.05	Compliance Milestone Type	A designator and associated metadata used to identify a type of compliance milestone.	(new)
CR.06	Compliance History	Monthly or quarterly compliance status information in relation to a regulatory requirement.	CR.04

⁵ The Affiliation data block was referred to as “Responsibility” in CRM Version 1.

Data Block ID	Data Block Name	Description	CRM I Data Block ID
E.01	Enforcement Identity	Basic identification information used for an enforcement action.	E.02
E.02	Enforcement Description	Information used to describe an enforcement.	E.03
E.03	Penalty Identity	The monetary amount that a facility is required to pay as a result of the enforcement. This penalty may consist of civil, stipulation, etc.	E.04
E.04	Enforcement Action Injunctive Relief	Information about any injunctive relief sought through, and/or required pursuant to, an Enforcement Action, but not including penalties, cost recovery, and Supplemental Environmental Project (SEP) obligations. Penalties, cost recovery and SEPs, are addressed in separate categories.	(new)
E.05	Applicable Environmental Citation ⁶	Local, State, or Federal environmental regulation, such as 40 CFR.	TBC.05
E.06	Enforcement Corrective Action	The activities that the facilities are required to perform as a result of the enforcement.	E.05
E.07	Enforcement Result	The outcome of a specified enforcement. This may include monetary amount and/or corrective action.	E.06
E.08	Enforcement Cost/Recovery	The monetary amount that may have been incurred by the issuing agency as a result of the enforcement activity.	E.07
E.09	Cost/Recovery Identity	Basic identification information used for an enforcement cost/recovery associated with an enforcement action.	E.08
E.10	Corrective Action Identity	Basic identification information used for a corrective action. Examples may include the purpose, corrective action name, issuing agency, etc.	E.10
E.11	Enforcement Action History	Information pertaining to ongoing enforcement actions. This includes appeal filed date, penalty collected date, final order issued date, etc.	E.11
EAE.01	Environmental Accident Identity	Basic identification information used for describing an environmental accident. Examples may include cause of accident, method of detection, etc.	EAE.01
EAE.02	Environmental Accident Detail	Detailed information for a given environmental accident. Examples may include address, substance, monitoring location, release quantity, etc.	EAE.02
EAE.03	Environmental Accident Reporter	Information about the person who reported the environmental accident event. Examples may include name of the reporter, reporter's telephone number, etc.	EAE.03
ESAR.01	Measure	Identifies the value and the associated units of measure for measuring the observation or analytical result value.	(new)
ESAR.02	Measure Unit Code	A code and associated metadata used to identify a unit of measurement.	(new)
ESAR.03	Result Qualifier Code	A code and associated metadata used to identify any qualifying issues that affect results.	(new)
ESAR.04	Laboratory Identity	Basic identification information for an entity or person responsible for analysis of substances.	(new)
ESAR.05	Accreditation	Information regarding the accreditation of a facility, laboratory or organization	(new)
ESAR.06	Monitoring Location Identity	Basic identification information for the location/site that is monitored or used for sampling.	(new)

⁶ E.05 was labeled as “To Be Categorized” in CRM Version 1.0, but later moved to the Enforcement Major Data Group.

Data Block ID	Data Block Name	Description	CRM I Data Block ID
ESAR.07	Manufacturing Equipment Identity ⁷	Detailed identification information used to describe equipment used in a process. Examples may include equipment ID, name, equipment size, usage type (e.g., storage tank, process vessel, boiler, etc.), equipment type (e.g., tank, distillation column, agitator, etc.), equipment description, maker, year installed, etc.	P.15
EN.01	Notice Identity	Basic identification information used in an environmental notice/advisory. Examples may include issuing agency, notice description, effective date, and expiration date.	EN.01
F.01	Facility Site Identity	Basic information used by an organization to identify a facility or site.	F.02
F.02	Facility Site Type	A designator and associated metadata used to represents the type of site a facility occupies.	(new)
F.03	Facility SIC	The Standard Industrial Classification (SIC), or type of business activity, occurring at the facility site.	(new)
F.04	Facility NAICS	The North American Industry Classification System (NAICS) code, or type of industrial activity, occurring at the facility site.	(new)
F.05	SIC Identity	The SIC code set used to classify the economic activities of most of the industries or kinds of business establishments in our economy. The SIC classification system defines economic activity into 10 divisions. These are further broken down into numeric codes that define major groups, industrial groups, and industries. Descriptive text is also available to define industrial subdivisions.	(new)
F.06	NAICS Identity	The NAICS code set used to classify the economic activities of business establishments, grouped hierarchically into economic sectors, economic subsectors, industry groups, and industries. Industry classifications are further subdivided into national classifications that are specific to the needs of each country.	(new)
F.07	Agency Identity ⁸	A designator and associated metadata used to identify a federal, state, or local agency.	TBC.06
F.08	Agency Type	A designator and associated metadata used to identify the type of federal, state, or local agency.	(new)
F.09	Facility Management Type	A designator and associated metadata used to identify the type of operation management currently at a facility.	(new)
F.10	District	Information about a division of territory, or a defined portion of a state, town, or city, made for administrative, electoral, or other purposes, such as a congressional district, judicial district, land district, school district, etc. Examples may include the Congressional District Number, which represents a Congressional District for a state, or the Legislative District Number, which represents a Legislative District within a state.	F.08
F.11	Alternate Name	Historic, secondary, or program-specific names used by the facility. This may include names such as "Doing Business As (DBA)".	F.09
F.12	Facility History	Historical tracking of facility data. This may include former identification information, name, location, etc.	F.18

⁷ Manufacturing Equipment was categorized under "Permit" in CRM I but is now categorized in the ESAR Major Data Group.

⁸ Agency Identity was in the "To Be Categorized" group, but is now with the Facility Major Data Group.

Data Block ID	Data Block Name	Description	CRM I Data Block ID
G.01	Grant	Grant information such as payment order, award, proposal, deliverables, and area of application, among others. Scope of administration is the high-level data that States and US EPA exchange.	G.01
G.02	Fund Source	Information about the origin of financial dispensation, such as the entity providing grant money.	G.02
G.03	Agreement and Memorandum of Understanding	Information related to an Agreement and Memorandum of Understanding.	G.03
L.01	License Identification	The name assigned to the license by a license issuing/granting organization to identify a license or license application. This may also include the alphanumeric identifiers assigned and used to issue the license by a license issuing/granting organization, as well as other alphanumeric identifiers used to identify a license or license application, and a brief description of the other license number/identifier context.	L.01
L.02	License Identity	Basic identification information for a license, including the State's unique license identifier/number, issuing agency, name, type, and description. A License differs from a Permit since it is administrated under a separate procedure (e.g., treatment operator license, pesticide application operator license, etc.) and does not contain monitoring requirements.	L.02
L.03	License Description	The type of license issued or granted to a facility or individual, represented by a License Identifier Data Block. Potential information may include a description of the specific professional practice or activity licensed (e.g., well-drilling, asbestos removal, stack testing, etc.)	L.03
L.04	Licensing Requirement	Description of activities, education, and certification required to obtain a license. This may include continuing education requirements needed to demonstrate competence in the licensed activity.	L.04
L.05	Money	The monetary amount associated with the specified product, action, program, findings, etc. Examples may include the purpose of monetary payment type (penalty, fees), amount, payment-due date, payment receipt date, and interest payment.	L.05
MA.01	Analysis Result	Information relating to the analysis of samples and "in-situ" measurements that are recorded and attached to the visit/sample to which they relate. For example, STORET stores metadata concerning the quality of results related directly to the results. A description of the analysis procedure together with actual sampling techniques deployed and the final results of the analysis are provided in a report.	MA.03
MA.02	Sample Identification	Information gathered through observing the environment during site visit(s). This may include descriptions of the sample collection process outlining the field work procedures and keyed to the specific location at which the field work is conducted (e.g., linking water and/or air quality measurements). Samples are described according to their medium and the intent for which they were collected while sample collection is documented by links between a sample and lists of methods and equipment. Samples can be created from other samples, by compositing, splitting, or sub-sampling. Field monitoring activities may be water, air, or sediment sample collection, biological specimen catch/trap events, or any measurements or observations obtained at the site.	MA.04
MA.03	Quality Assurance Identity	Information about QA/QC steps taken for sampling. Examples may include blank, duplicate, and spike samples.	MA.05

Data Block ID	Data Block Name	Description	CRM I Data Block ID
MC.01	Inspection	Information related to an official visit to the regulated entity on a periodic basis for compliance purpose.	MC.03
MC.02	Inspection Identity	Basic identification information about an official review. Examples may include inspection type, inspection start date, inspection end date, inspector, facility contact, and the State's unique inspection identifier/number.	MC.04
P.01	Permit Identity	Basic identification information for a permit. Examples may include the State's unique permit identifier/number, permit name, permit type, permittee contact, permit feature type (e.g., Certificate Of Coverage, Notice of Coverage, etc), permitting dates and issuing agency.	P.01
P.02	Permitted Feature	Information about the permitted feature of a permit. A permitted feature is a unit, physical structure, feature, or process described in a permit.	(new)
P.03	Permit Administration	Administrative information about the permit.	(new)
P.04	Permit Limit Condition	Regulatory limit for a given release that is allowable under a permit. It also contains requirement information used to describe the conditions for which a release limit will apply.	P.12
P.05	Monitoring Condition	Administrative information that describes the monitoring requirements.	P.04
P.06	Permit Type	A code and associated metadata used to identify the type of permit issued or granted to a regulated entity.	(new)
P.07	Permit Event	Identifies dates associated with a permit or permit application process.	(new)
P.08	Record-Keeping Requirement	Record-keeping requirements as required by a given permit or regulatory statute (e.g., General Permit). Examples may include record retention time, contents of record, etc.	P.11
P.09	Process Specification	Comprehensive information about a process. This information may contain process identity, manufacturing equipment, control equipment/device, operating scenarios, production description, etc.	P.13
P.10	Process Identity	Basic identification information used to describe a process. A process is a series of manufacturing procedures to produce a finished product or an intermediate. Examples may include process name, process identifier/number, batch or continuous, expected yield, and operation description.	P.14
P.11	Process Activity	Detailed information that describes specific operation and/or activity performed for a process. This information may include manufacturing equipment used, control equipment, activity description (e.g., heating, distillation, drying, sweep, etc.), time used, substance involved, etc.	P.16
P.12	Operation Scenario	A manufacturing process could be run under different operating scenarios. Examples include process run without control equipment, process run using alternate substances, process run using alternate procedures, etc. Different operating scenarios will result in different emissions and be subject to different regulatory requirements.	P.17
P.13	Substance Consumption	Amount of substance used during a specified process. Examples include substance name, start date, end date, consumption amount, and unit.	P.18
P.14	Production Description	Information used to describe production information for calculating actual air emissions. Examples may include process batches per year, date that the process is run, gallons of fuel used, material throughput for a storage tank, etc.	P.19

Data Block ID	Data Block Name	Description	CRM I Data Block ID
R.01	Release Classification	Information used to denote environmental releases. Examples may include on-site release, off-site release, release class description, in-state release, out-of-state release, etc.	R.01
R.02	Release Quantity	Information related to observed substance release. This may include substance released, released quantity, unit, released date and time.	R.02
R.04	Release Detail	Specific information about the quantity of each release. Examples may include substance name, release quantity, release date, unit, release nature (e.g., potential-to-emit or actual emissions), etc.	R.04
R.07	Receiving Environment	Environmental media (e.g., air, soil, groundwater, navigable water, name of receiving environment) into which substances may be discharged or emitted. Examples may include media type, receiving environment name, receiving environment description.	R.07
R.08	Waste Management/Disposal	Information about disposal and management of waste. This includes waste disposal type, waste treatment, etc.	R.08
R.09	Emission Point	Information used to identify a "point" where a regulatory limit could apply. An emission point could be a physical vent stack (e.g., from an incinerator, a process vessel, etc.) or a virtual boundary from process areas (e.g., fugitive emissions from Factory 29). Examples may include vent diameter, stack orientation, elevation, and distance to the property line.	R.09
R.10	Waste Disposal Activity	Information used to describe how wastes are disposed of. Examples may include disposal type (e.g., landfill, recycle, underground injection, transfer, energy recovery), waste transporters, final disposal facility, etc.	R.10
R.11	Waste Treatment Activity	Information used to describe how wastes are treated. Waste treatment normally occurs before final disposal. Examples may include treatment method, treatment efficiency, reduction activity, etc.	R.11
RMF.01	Reference Method	Standard testing methods approved for regulatory compliance or environmental monitoring purposes. Methods are specific to the substance being analyzed, analysis instrumentation used, and the environmental media or matrix from which the sample is taken.	RMF.01
RMF.02	Reference Factor	Standard quantity which is multiplied by another quantity (e.g., emission factors used to calculate air emissions) published by regulatory agencies for the purpose of estimating releases for potential-to-emit or actual emissions.	RMF.02
RPT.01	Report Identity	Basic identification information used for an official account or statement.	TBC.02
RPT.02	Report Type Code	A code and associated metadata used to identify a type of report.	(new)
RPT.03	Reporting Condition	Data relating to a series of required reports.	(new)
RPT.04	Form Identity	Basic identification information used to describe a paper or electronic document with blanks for the insertion of details or information.	(new)
RPT.05	Form Instruction	Detailed instructions on filling out forms.	(new)
RPT.06	Submission Header	General description of contents and purposes in a report or transmission file. Examples may include the sender, the receiver, contact information, report type, and etc.	TBC.03
RPT.07	Certification	Information required as part of a regulatory submission, permit application, inspection report, etc. Examples may contain Data Blocks such as Certification Description, Contact, Telephone, Electronic Address, etc.	TBC.08
RPT.08	Metadata	General information about the file format definition used in the selected transmission. Examples may include schema name, schema version, and schema's author.	TBC.04

Data Block ID	Data Block Name	Description	CRM I Data Block ID
RPT.09	Certification Description	Information used for validating the authenticity of an activity. Examples may include the certification statement and certification date.	TBC.09
S.01	Substance Identification	Basic identification information used in the discovery of a material that is used or measured.	S.01
S.02	Chemical Substance Identity	The Chemical Identification Block provides for the use of common identifiers for all chemical substances regulated or monitored by environmental programs.	(new)
S.03	Biological Substance Identity	The Biological Identification Block provides for the use of common identifiers for all biological organisms regulated or monitored by environmental programs.	(new)
S.04	Substance Property	Chemical or physical properties of a given substance or biological entity. Examples may include molecular weight, melting point, Antoine's Constants, activity coefficient, molecular structure, functional behavior, toxicological impact, etc.	S.04
S.05	Regulatory List	A list of chemical groups defined as per regulatory citation and/or compliance evaluation purpose. Examples may include Hazardous Air Pollutants (HAPS), volatile organic chemical (VOC), total toxic organics (TTO), Pharmaceutical MACT chemicals, etc. This Data Block could be used with the Substance Identity Data Block to designate the groupings and/or regulatory requirements that the substance is subject to.	S.05
S.06	Substance Usage	Information about where and how a substance originated and has been used. Examples may include substance origin (e.g., imported, manufactured on-site, recycle by-product, treatment by-product, etc), usage description, and usage location.	S.07
SD.01	Geographical Location Description	Extensive list of geographic identifiers used to clearly mark an object's precise location. Examples may include latitude, longitude, etc.	SD.02
SD.02	Geographic Reference Point Code	A code and associated metadata used to identify a geographic reference point.	(new)
SD.03	Geographic Reference Datum	Information that describes the reference datum used in determining geographic coordinates.	(new)
SD.04	Coordinate Data Source Code	A code and associated metadata used to identify a data source of coordinate data.	(new)
SD.05	Geometric Type Code	A code and associated metadata used to identify a geometric entity represented by one point or a sequence of points.	(new)
SD.06	Area	Boundaries that could be naturally formed or determined by political, administrative, and other considerations. Examples may include area type (man made or natural), area description, and district.	SD.01
SR.01	Control Methodology	A process and/or tool used to manage storage, disposal, treatment, or other handling protocols designed for and/or used.	(new)
SR.02	Point Source Identification	Basic identification information used to describe stationary source of emissions confined to a single location. An example may include an electric power plant which can be identified by name and location.	SR.01
SR.03	Non-Point Source Identity	Basic identification information used to describe sources of emissions that are not confined to a single location. Examples may include agriculture, saltwater intrusion, etc.	SR.02

Data Block ID	Data Block Name	Description	CRM I Data Block ID
SR.04	Mobile Source Identity	Basic identification information used to describe non-stationary sources of emissions. Examples may include vehicle or equipment with mobile capability (e.g., small engines, ships, aircraft, etc.).	SR.03
SR.05	Source Identification	Basic identification information used in source discovery, including point, non-point, and mobile source.	SR.04
SR.06	Control Equipment Identity	Detailed identification information used for control equipment/device. Examples may include the make and model number.	SR.05
SR.07	Control Specification	Detailed information about the setup and/or operating conditions of control equipment when used to reduce the amount of pollutants being released into environment. Examples include condenser cooling media, cooling media temperature, incinerator burner temperature, acid scrubber flow rate, etc.	SR.06

4.3 Data Blocks Removed

In moving from CRM Version 1 to CRM Version 2.0, additional refinements to the model occurred. These refinements were a result of additional harmonization with the EDSC Data Standards as well as additional model analysis by CRM Workgroup members. As a result of the model enhancements, the following data blocks have been removed or superseded:

Table 3: Data Blocks Removed from CRM Version 2.0

CRM I Data Block ID	Data Block Name	Description	Reason for Removal
C.01	Contact Identification	Basic contact information which may include Data Blocks such as <i>Individual Identity</i> , <i>Facility Identity</i> , <i>Tribal Identity</i> , <i>Regulatory Entity</i> , and <i>Responsibility</i> .	Harmonization with EDSC standards
C.05	Address Identification	Information related to deliverable and non-deliverable mailing and location address. The deliverable address information may contain street address, supplemental address, city name, state name, country name, state code and zip code. The non-deliverable address information may contain only location descriptions.	Harmonization with EDSC standards – address now split into mailing and location address
CR.01	Violation Identification	Basic identification information used for defining a noncompliance with one or more legally enforceable obligations by a regulated entity, as determined by a responsible authority.	Harmonization with EDSC standards
E.01	Enforcement Identification	Basic identification information for enforcement, including the State's unique enforcement identifier/number, issuing agency, enforcement action date, type, action, and enforcement citation/reference.	Harmonization with EDSC standards
F.01	Facility Identification	Basic identification information for a facility site, including the US EPA or State facility registry identifier, geographic address, and geopolitical descriptors. May also include locality name, county or State FIPS codes, Tribal land name, geographic coordinates of latitude/longitude, etc.	Harmonization with EDSC standards

CRM I Data Block ID	Data Block Name	Description	Reason for Removal
F.07	Facility Activity	Information used to define the business activities occurring at the facility site. These business activities may be defined in terms of Standard Industrial Classifications (SIC)/ North American Industry Classification System (NAICS) codes.	Harmonization with EDSC standards – replaced with Facility SIC and Facility NAICS blocks
F.03	Environmental Interest	Information related to the environmental program interests associated with the facility site. Examples may include <i>Contact Identification, Address Identification, Facility Activity</i> , etc.	Harmonization with EDSC standards
P.03	Permit Identification	The name/information assigned to the permit by a permit issuing/granting organization to identify a permit or permit application. This information may contain the State's unique permit identifier/number, permitting date, permit type, etc.	Harmonization with EDSC standards
P.10	Monitoring Requirement	Information about a regulated substance that may be monitored. Examples may include sample frequency, sample type, monitoring conditions, etc.	Harmonization with EDSC standards
TBC.01	Report Identification	Identifying information used in a report or submission to satisfy the compliance requirements. This information may include reporting frequency, reporting date, etc.	CRM Workgroup II analysis
MA.01	Monitoring Location	The location where monitoring activities occurred. These locations may be specified and defined on the permit/license. These locations may also be used for ambient monitoring by agencies for purposes of reporting on the quality of their ambient air and water. Monitoring location may include information such as address, site type, site name, etc.	Replaced with ESAR.06
MA.02	Monitoring Location Identity	Basic identification information for the location/site that is monitored or used for sampling. Examples may include site identifier/number, sample location name, site name, site type, sampling group, and site group.	Replaced with ESAR.06
S.06	Substance Identity	Basic identification information for the classification system used to name a substance. Examples may include substance name, synonym, US EPA Chemical Registry System ID, Chemical Abstracts Service (CAS) number, regulatory reference, and the associated tracking system that supplies the identifier number.	Merged with S.01

4.4 Major Data Group Inventory Details

Data Blocks were identified from the environmental data found on the Exchange Network. When Data Blocks are logically grouped together, they form a specified business area (Major Data Group). Major Data Groups, individually or used together, will represent the flows of environmental data on the Exchange Network.

When a Data Block is first identified, it will take on its Major Data Group's identifying letter to indicate where it originated. However, this does not mean that the Data Block belongs to or can only be used in that Major Data Group.

For simplicity, only Data Blocks that have the same identifying letter as their residing Major Data Group will have their exemplary Data Element defined. For re-used Data Blocks that have a different identifying letter than the residing Major Data Group, their example Data Element will not be shown. Please refer back to the Major Data Group for its exemplary Data Element. As for the detailed description about each of the Data Blocks and Major Data Groups shown in diagrams 10 to 28, please refer back to chapter 4.

Currently, 20 Major Data Groups have been identified and used in the CRM. These 20 Major Data Groups were identified and defined in Table 1 of Section 3.4. The remainder of this section illustrates the Major Data Groups, each depicted with example data elements.

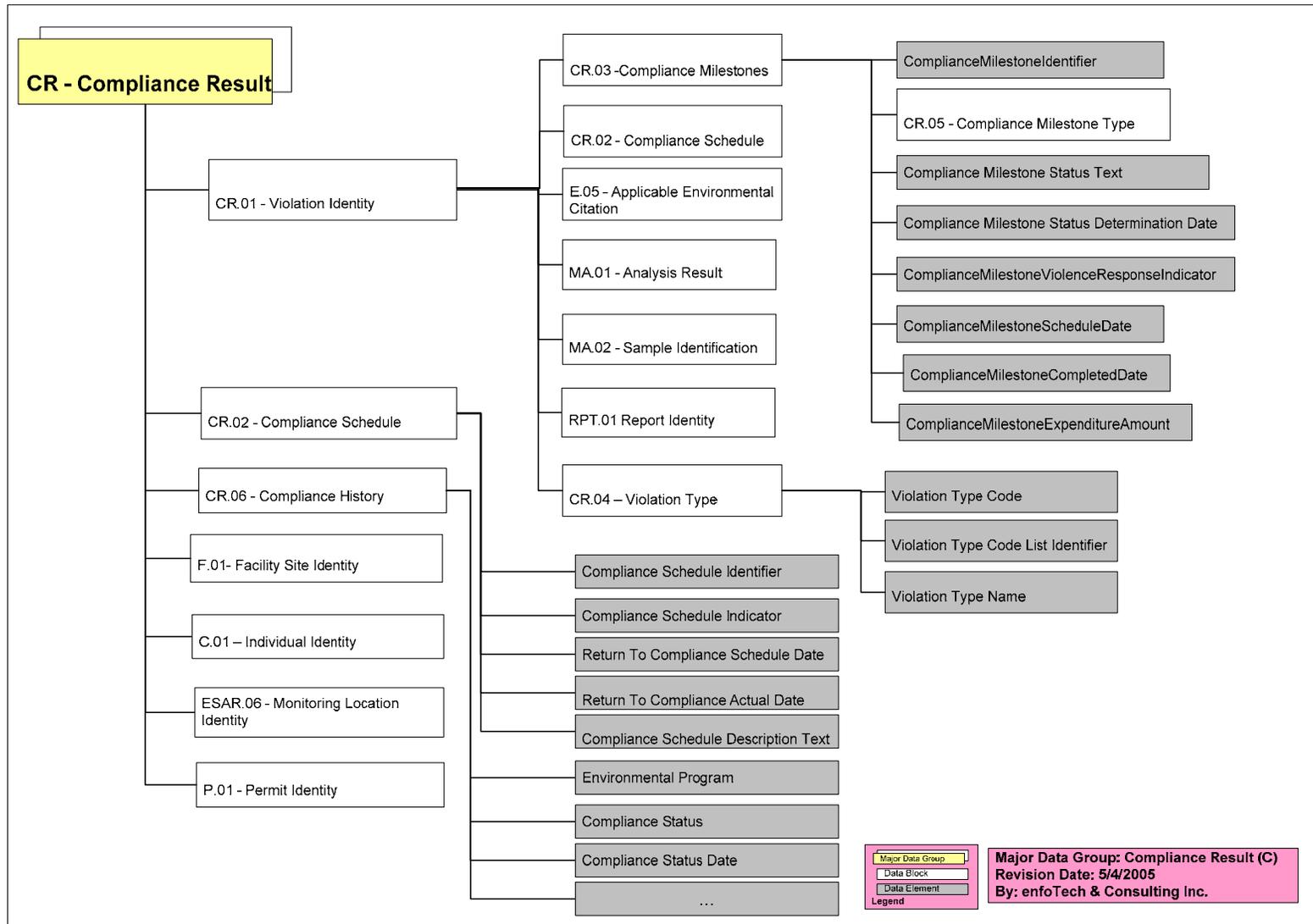


Diagram 8: Major Data Group of Compliance Result (CR)

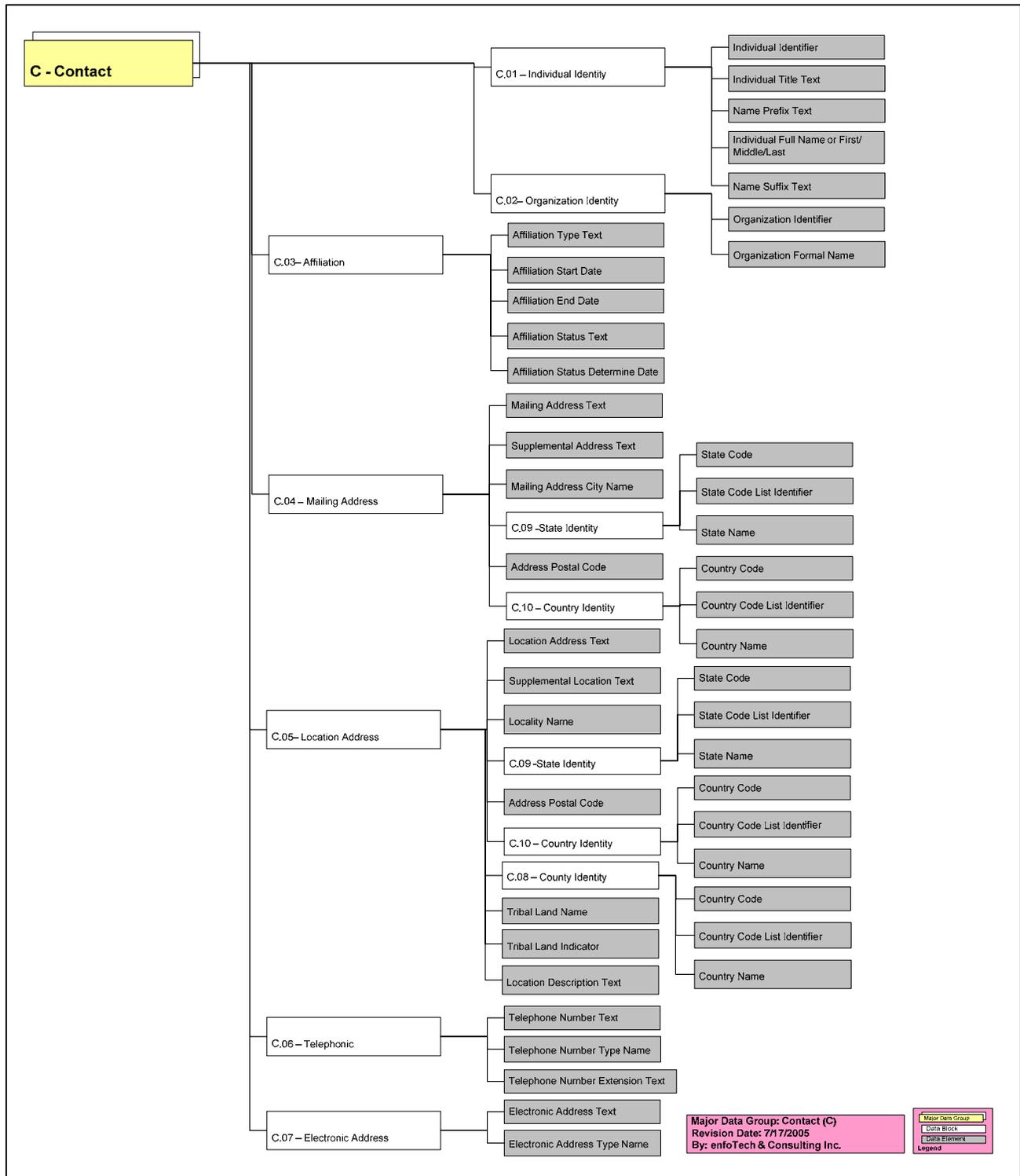


Diagram 9: Major Data Group of Contact (C)

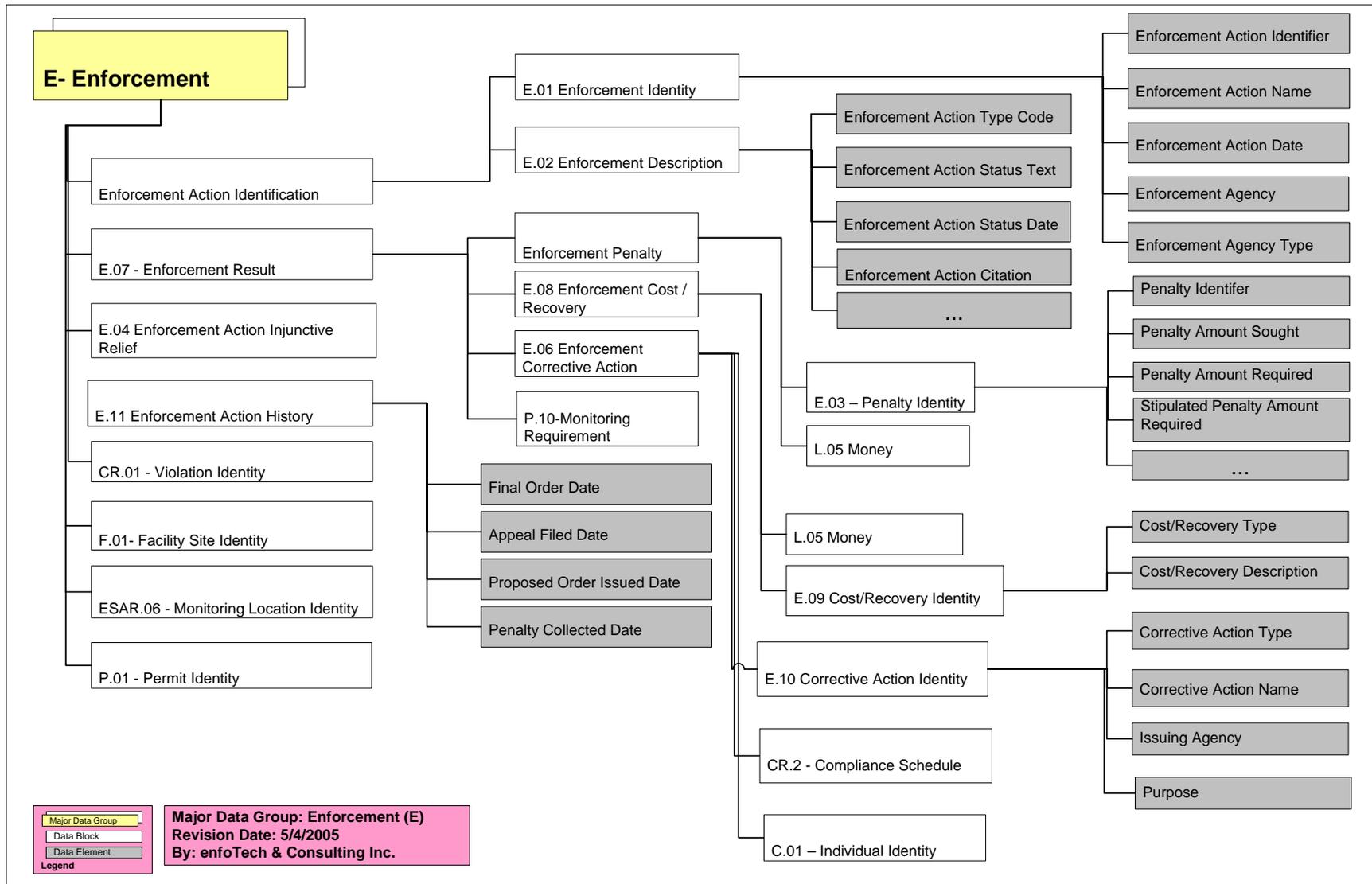


Diagram 10: Major Data Group of Enforcement (E)

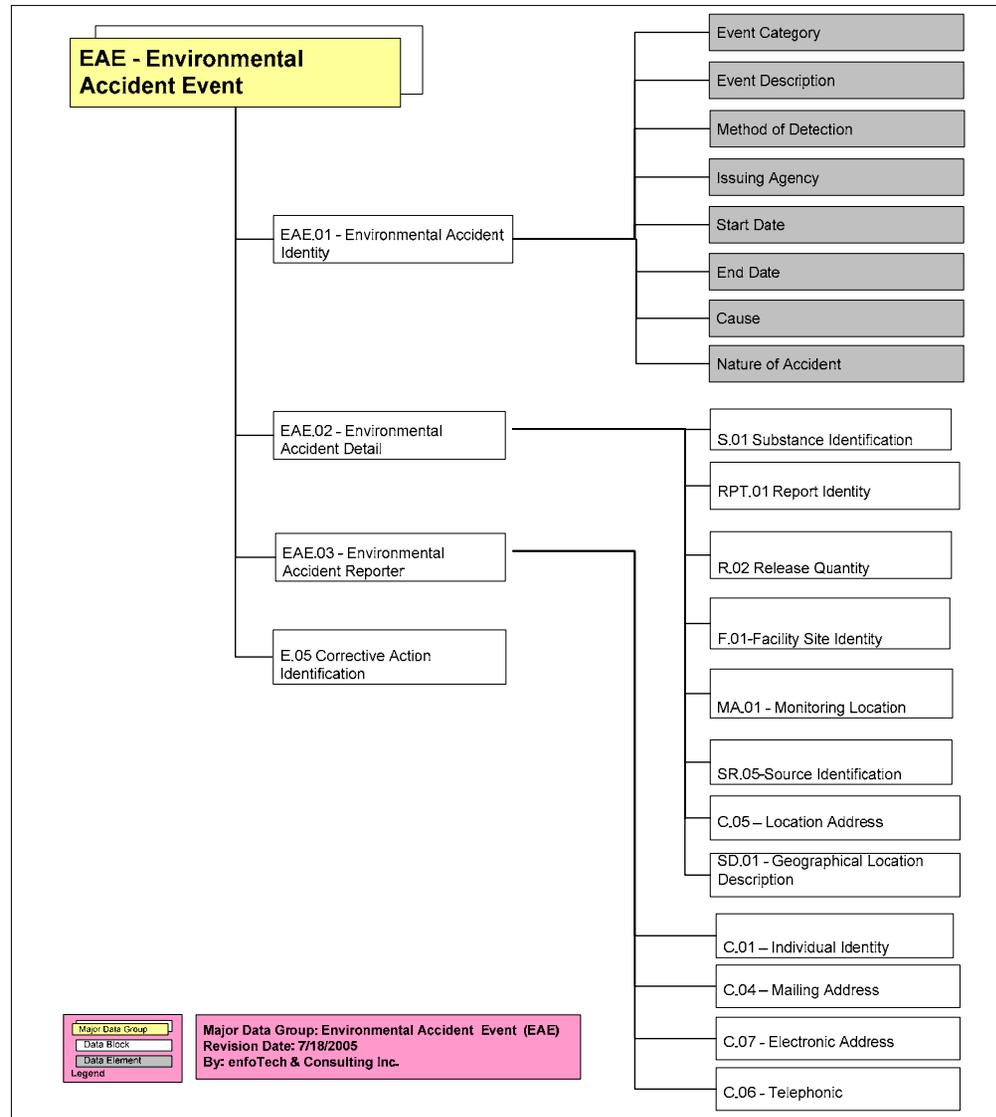


Diagram 11: Major Data Group of Environmental Accident Event (EAE)

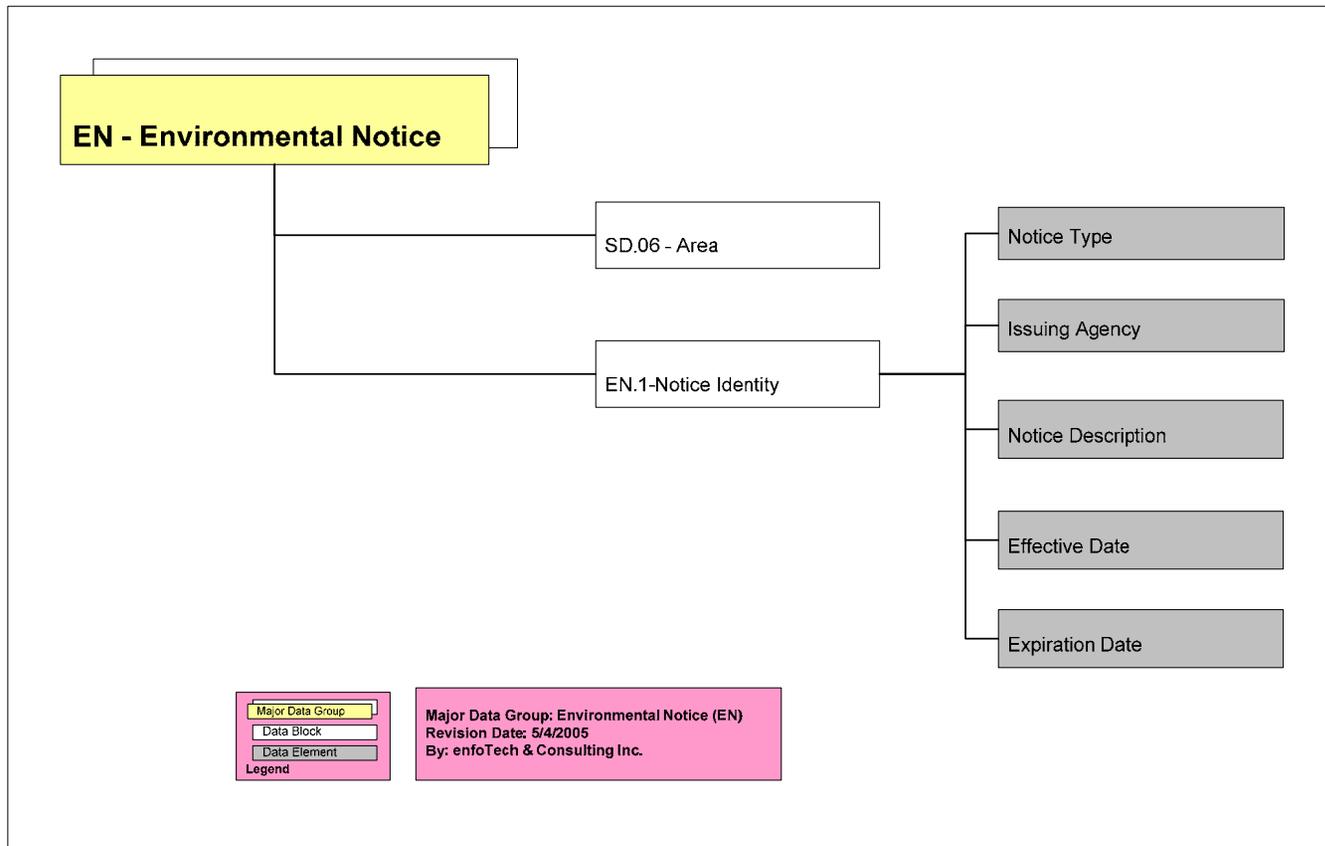


Diagram 12: Major Data Group of Environmental Notice (EN)

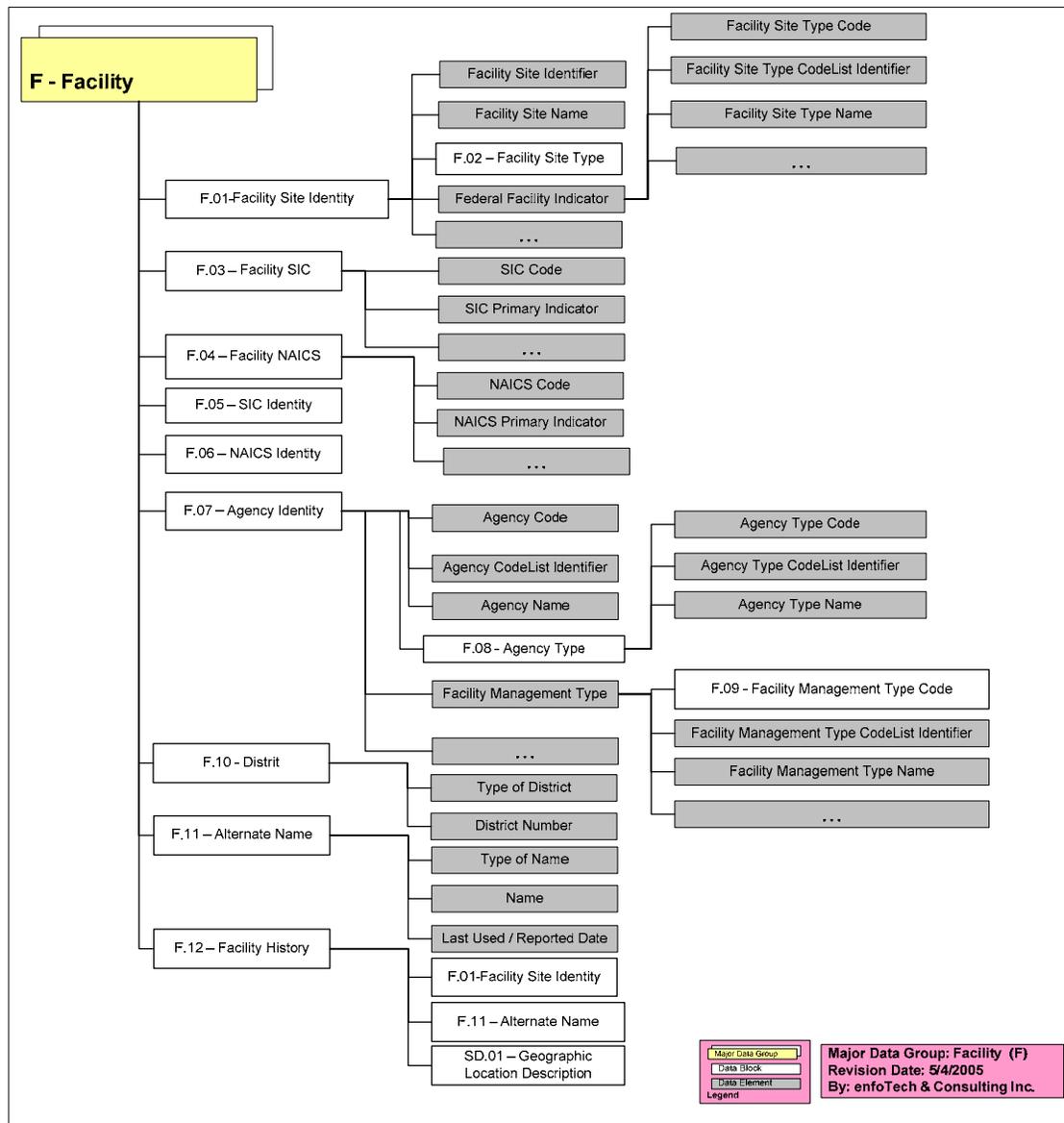


Diagram 13: Major Data Group of Facility (F)

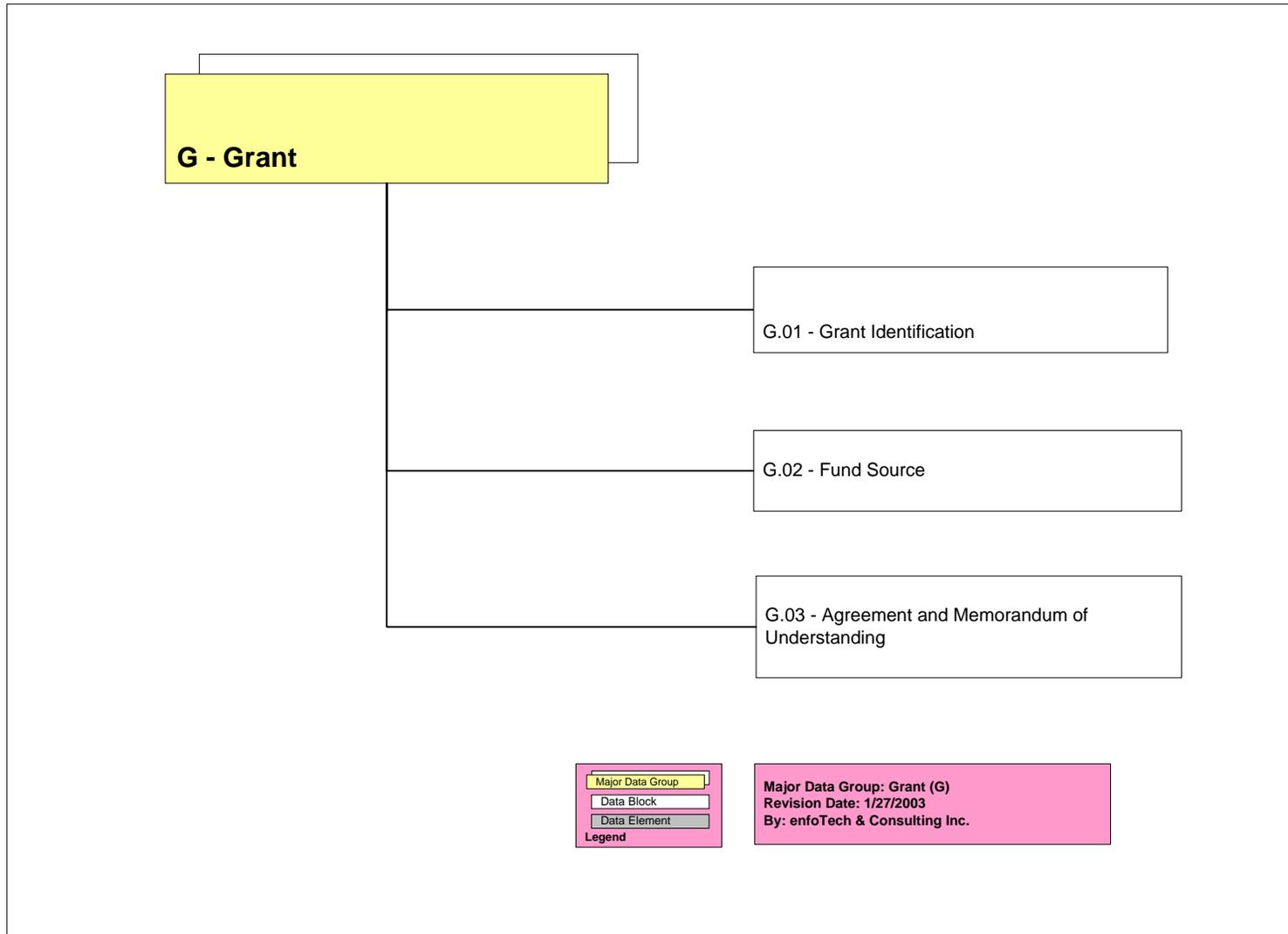


Diagram 14: Major Data Group of Grant (G)

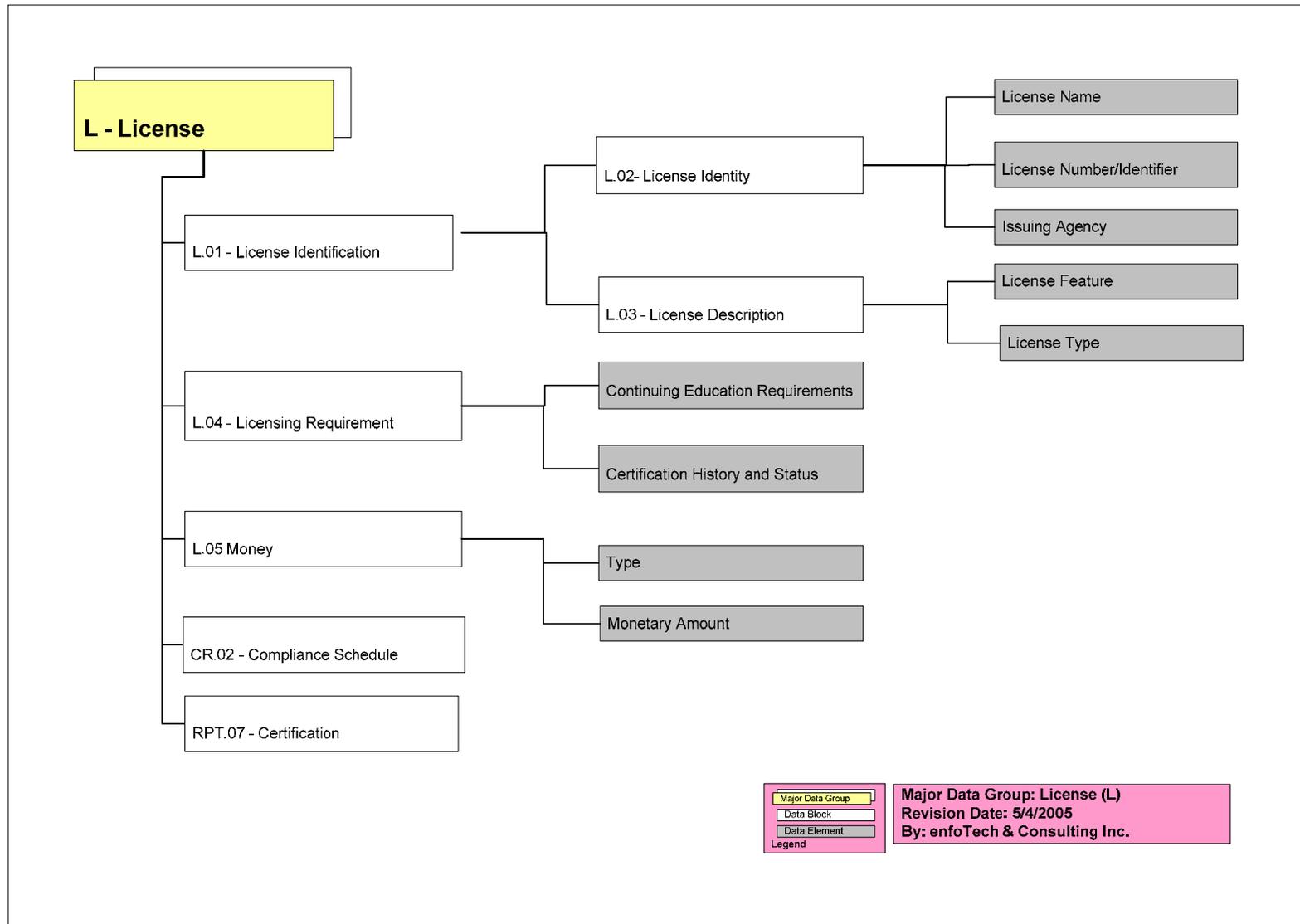


Diagram 15: Major Data Group of License (L)

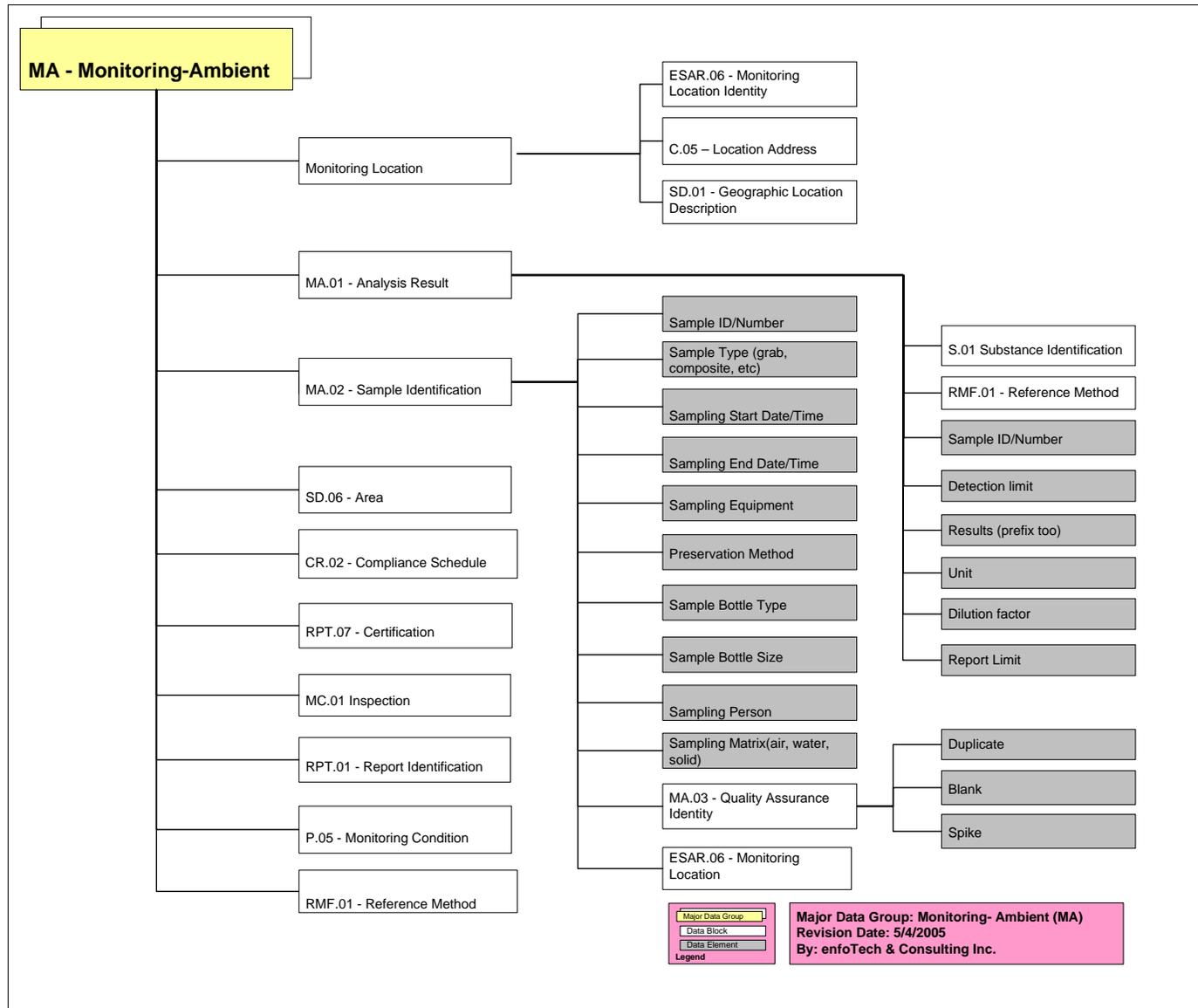


Diagram 16: Major Data Group of Monitoring Ambient (MA)

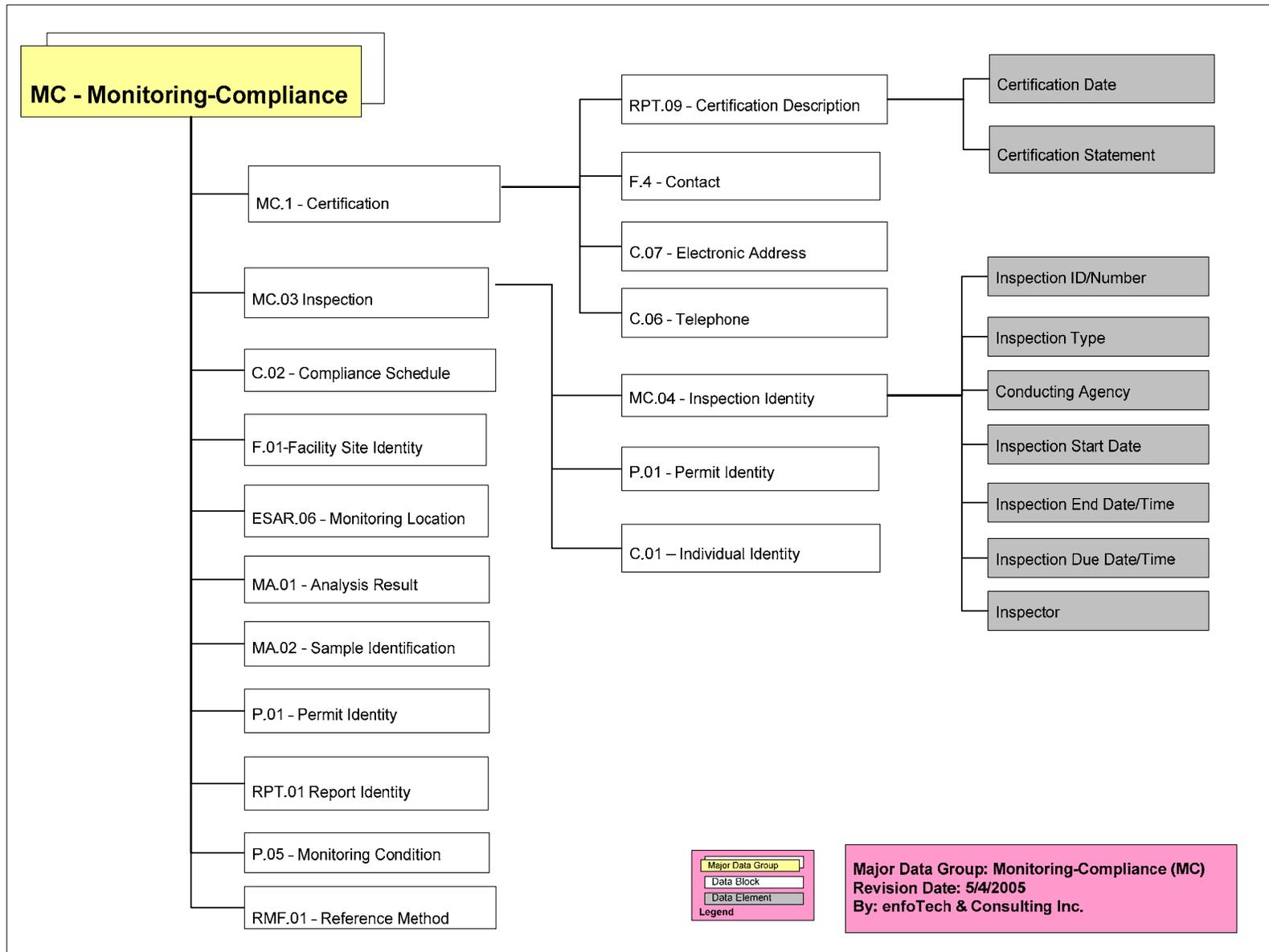


Diagram 17: Major Data Group of Monitoring Compliance (MC)

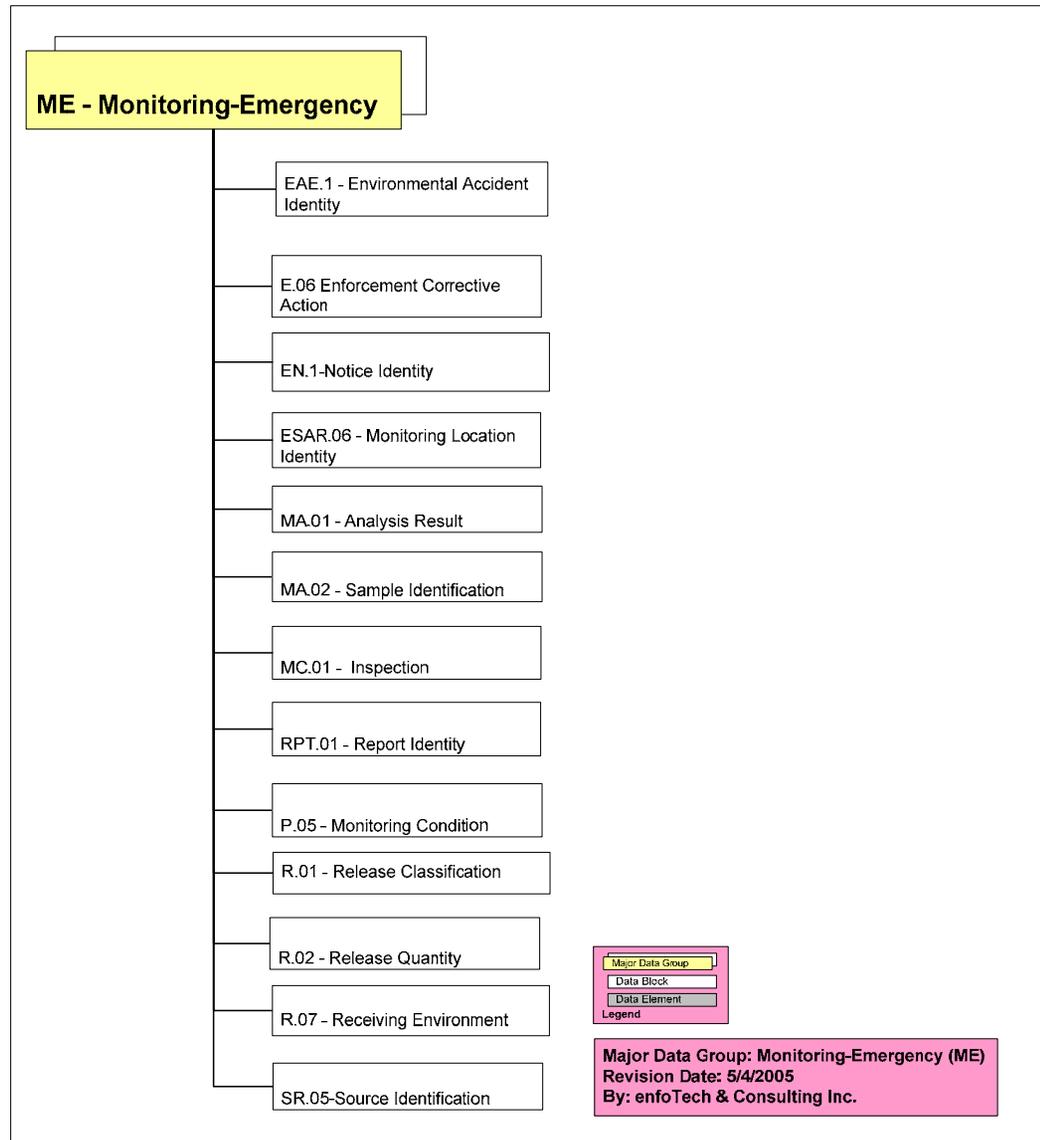


Diagram 18: Major Data Group of Monitoring Emergency (ME)

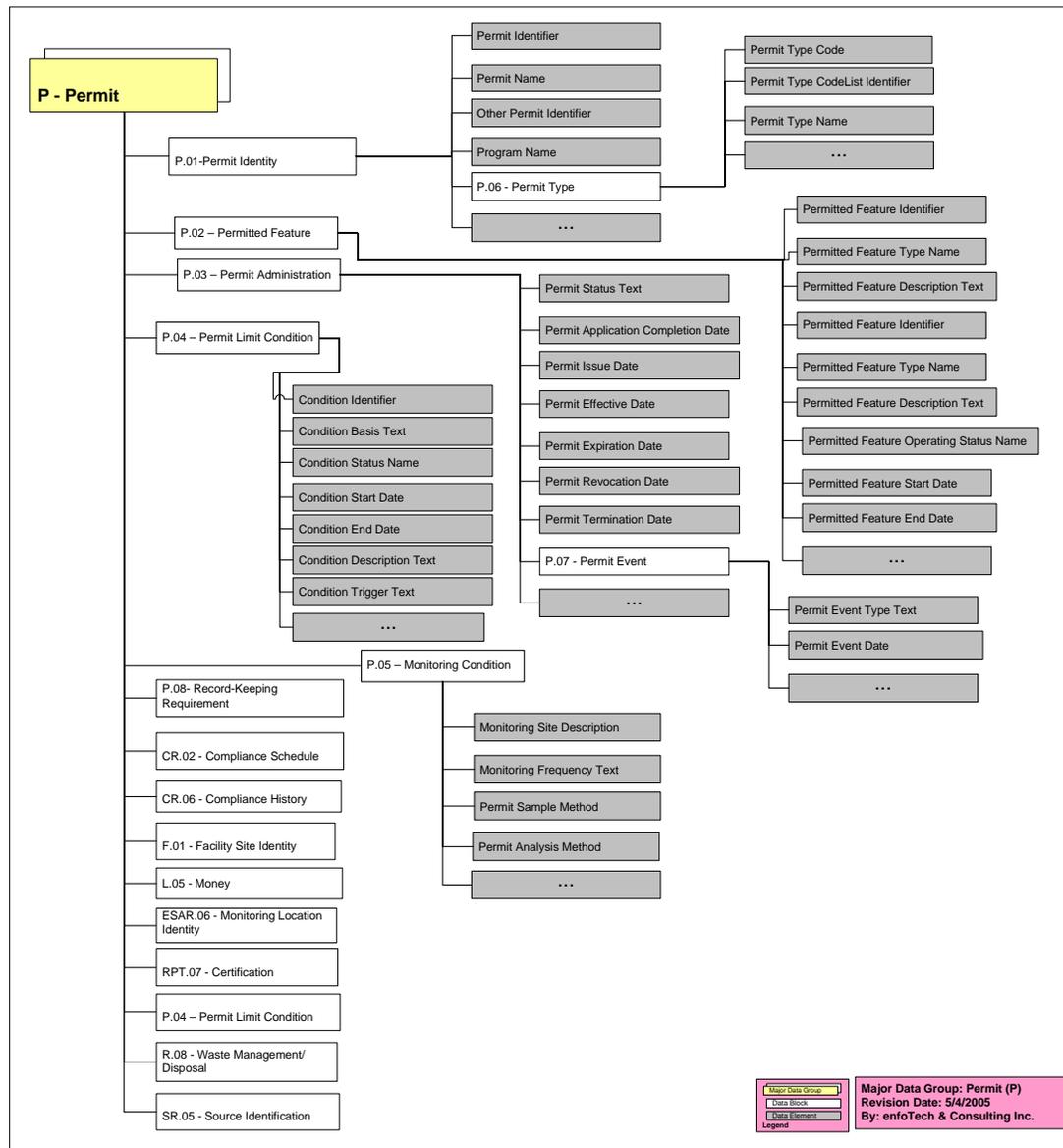


Diagram 19: Major Data Group of Permit (P)

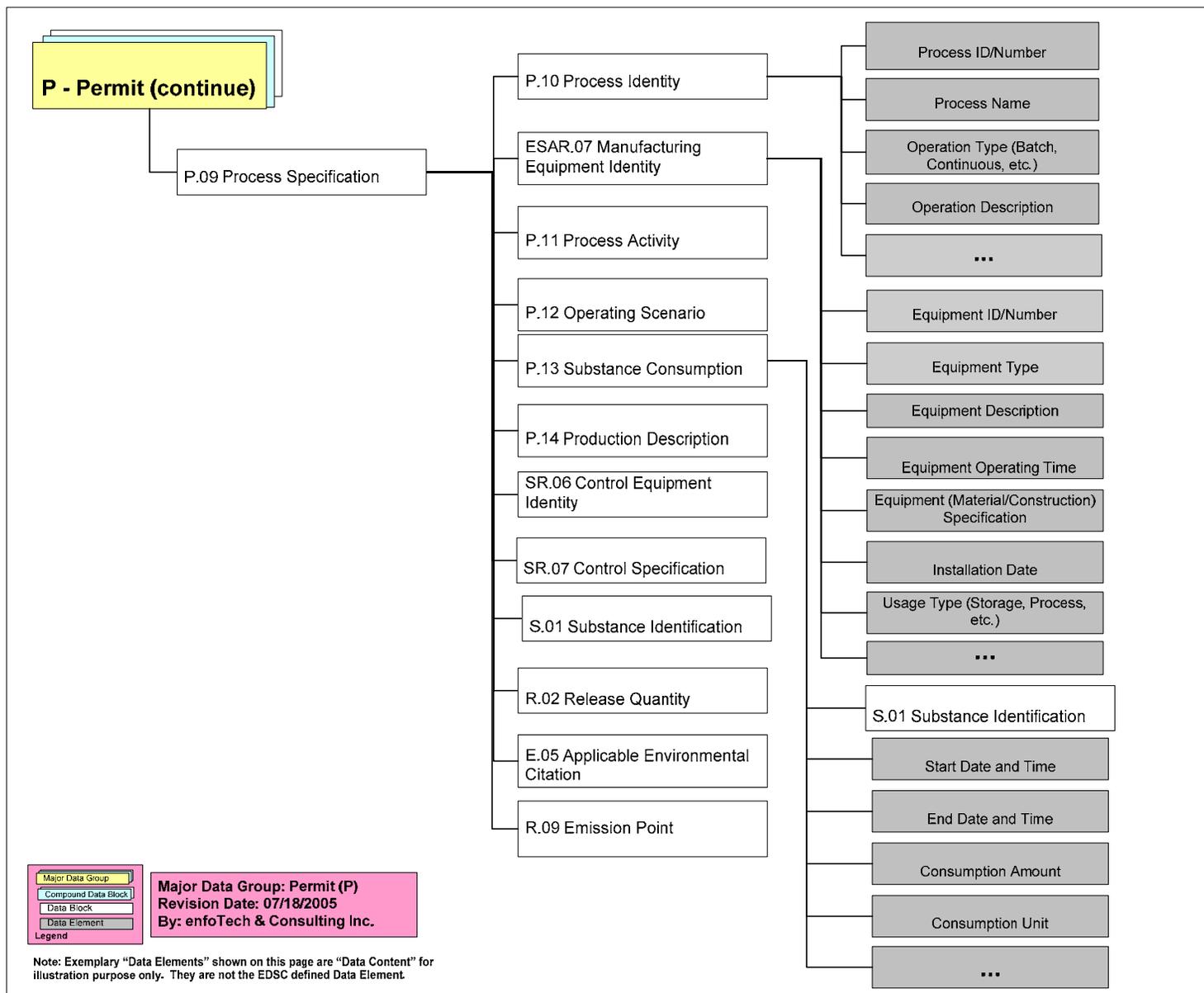


Diagram 20: Major Data Group of Permit (P) - Continued

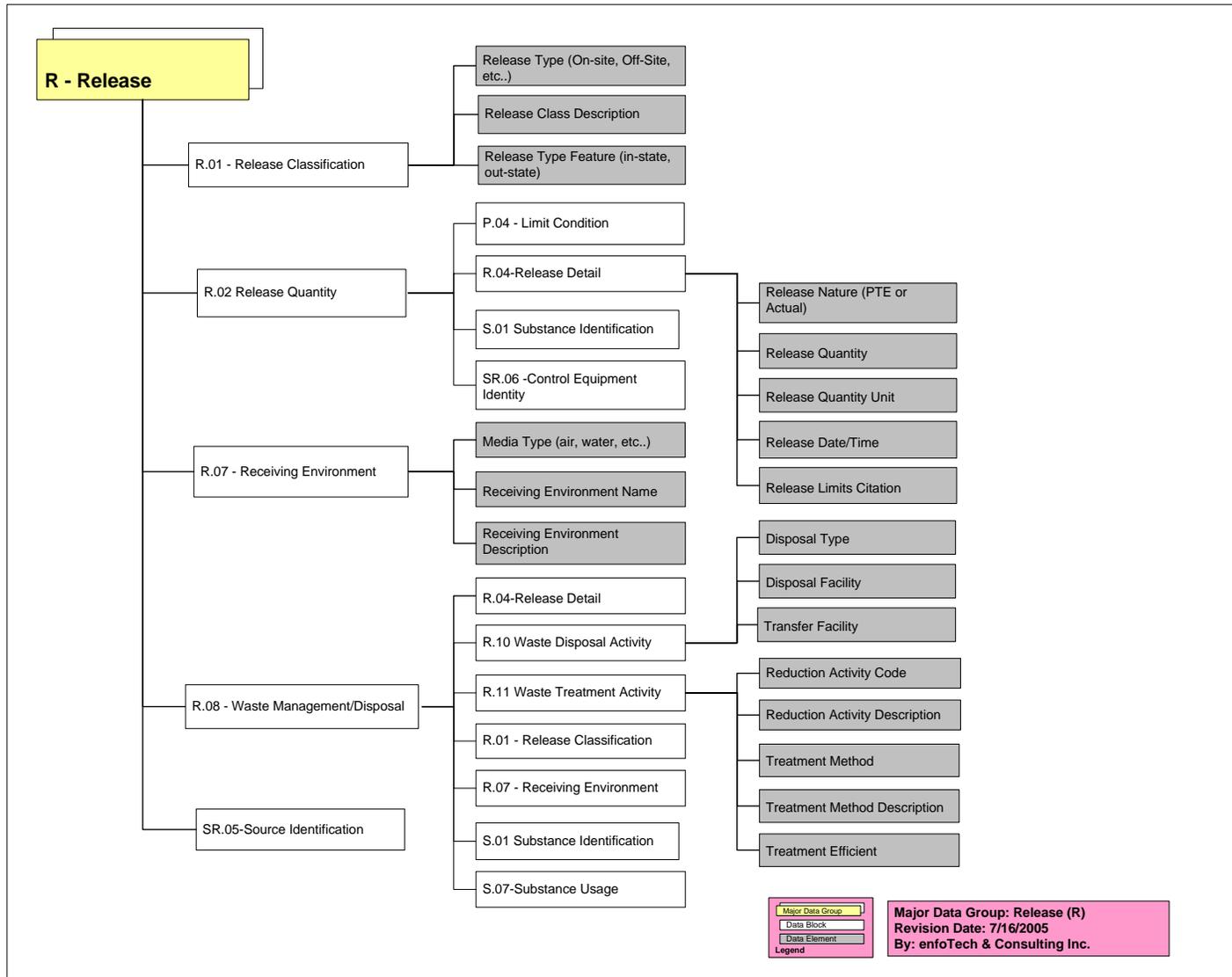


Diagram 21: Major Data Group of Release (R)

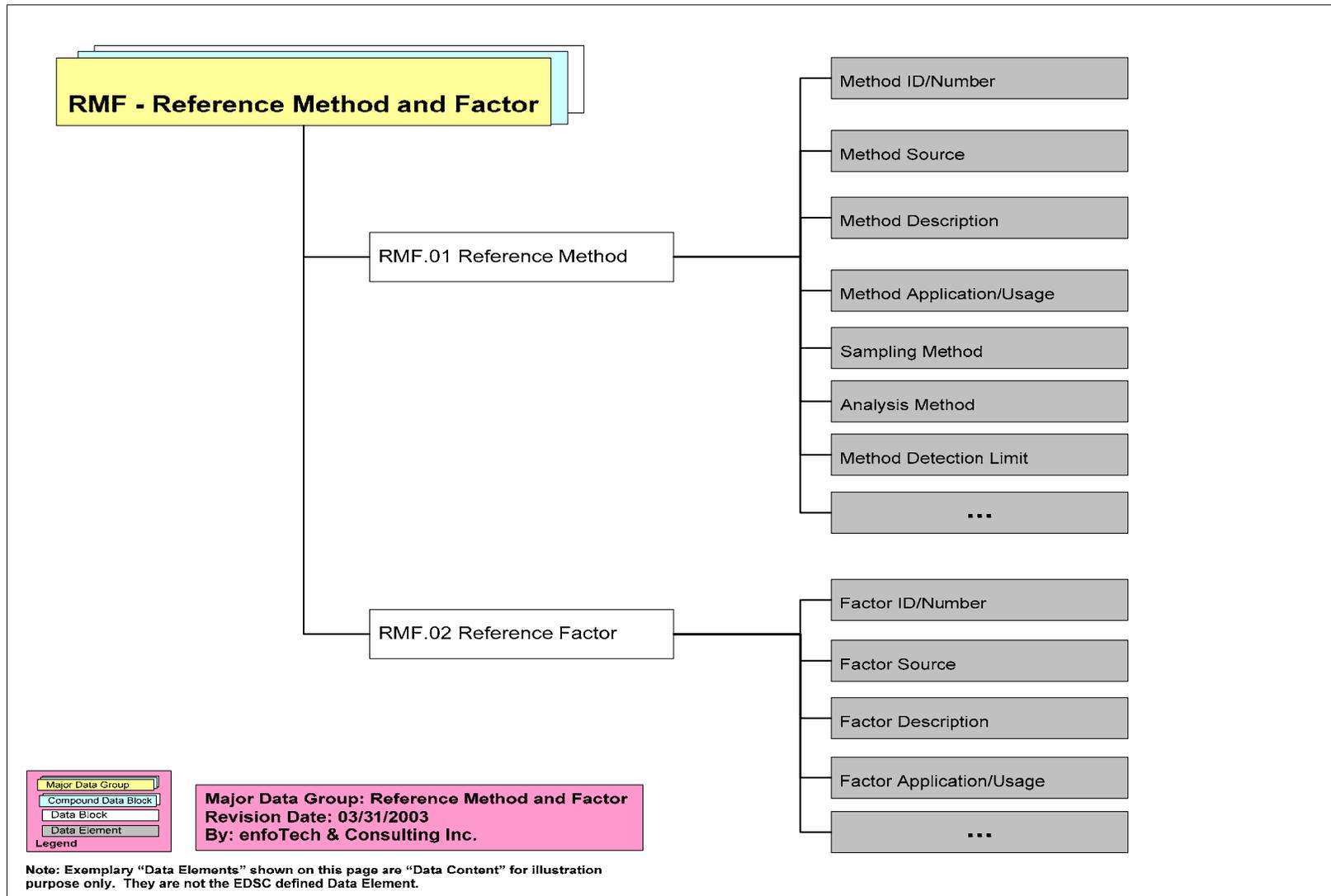


Diagram 22: Major Data Group of Reference Method and Factor (RMF)

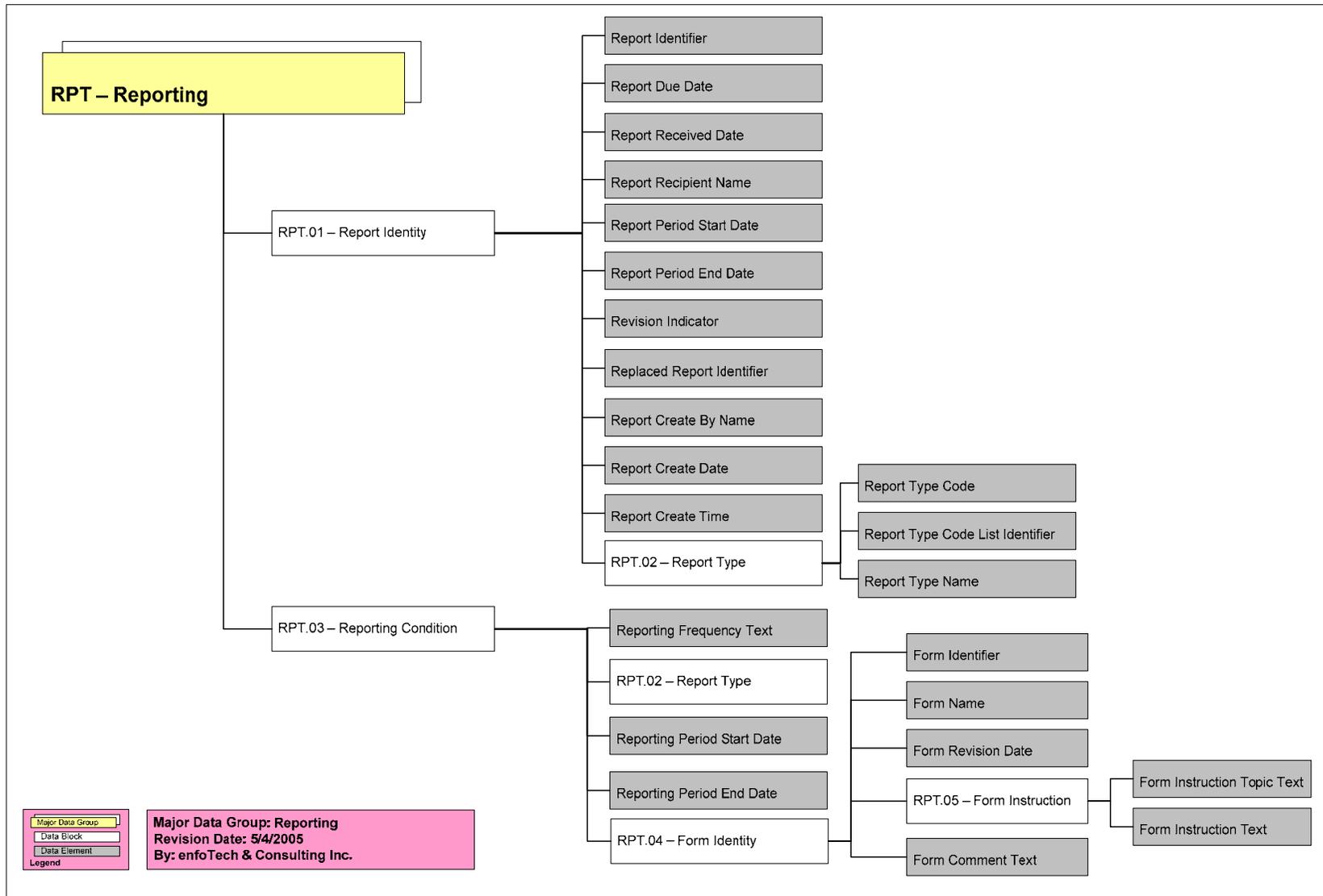


Diagram 23: Major Data Group of Reporting (RPT)

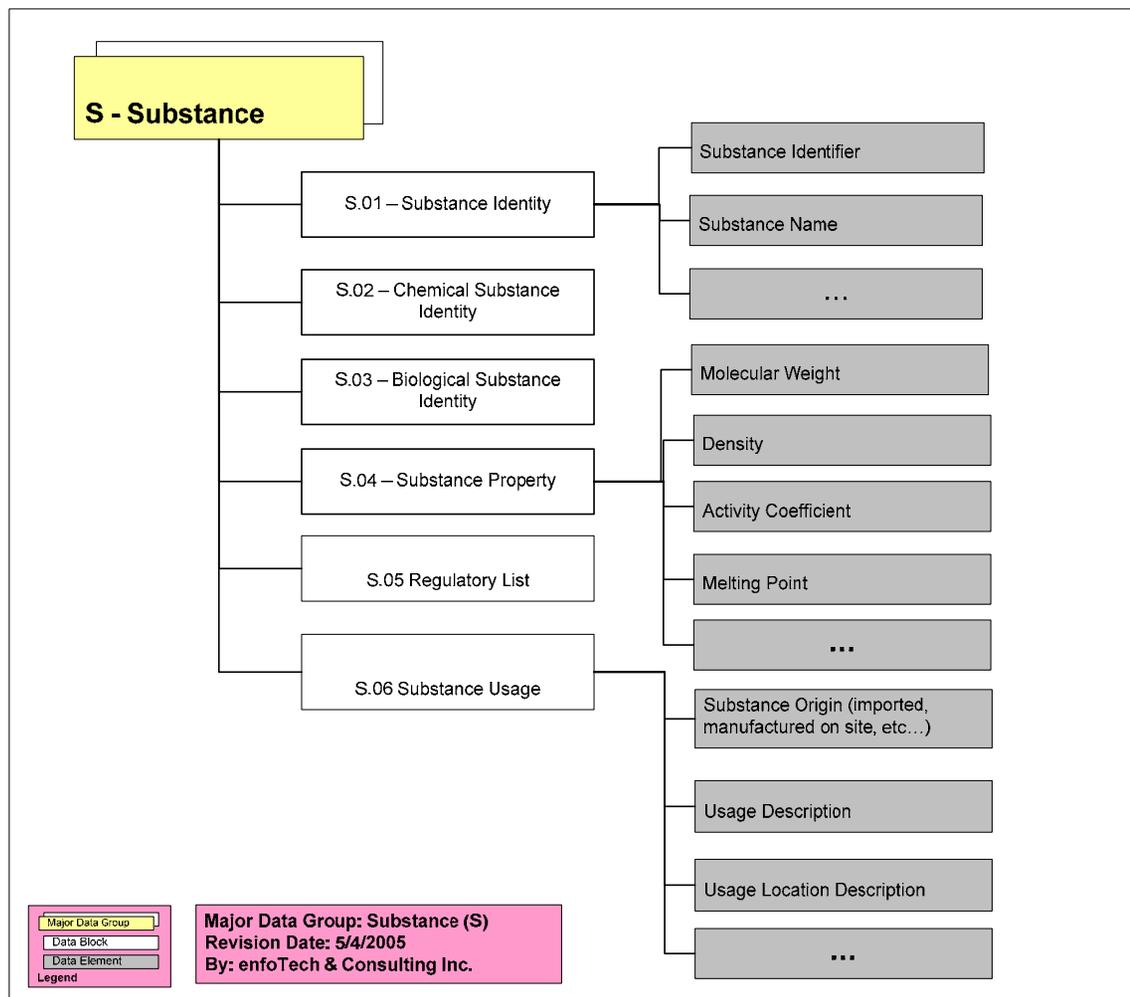


Diagram 24: Major Data Group of Substance (S)

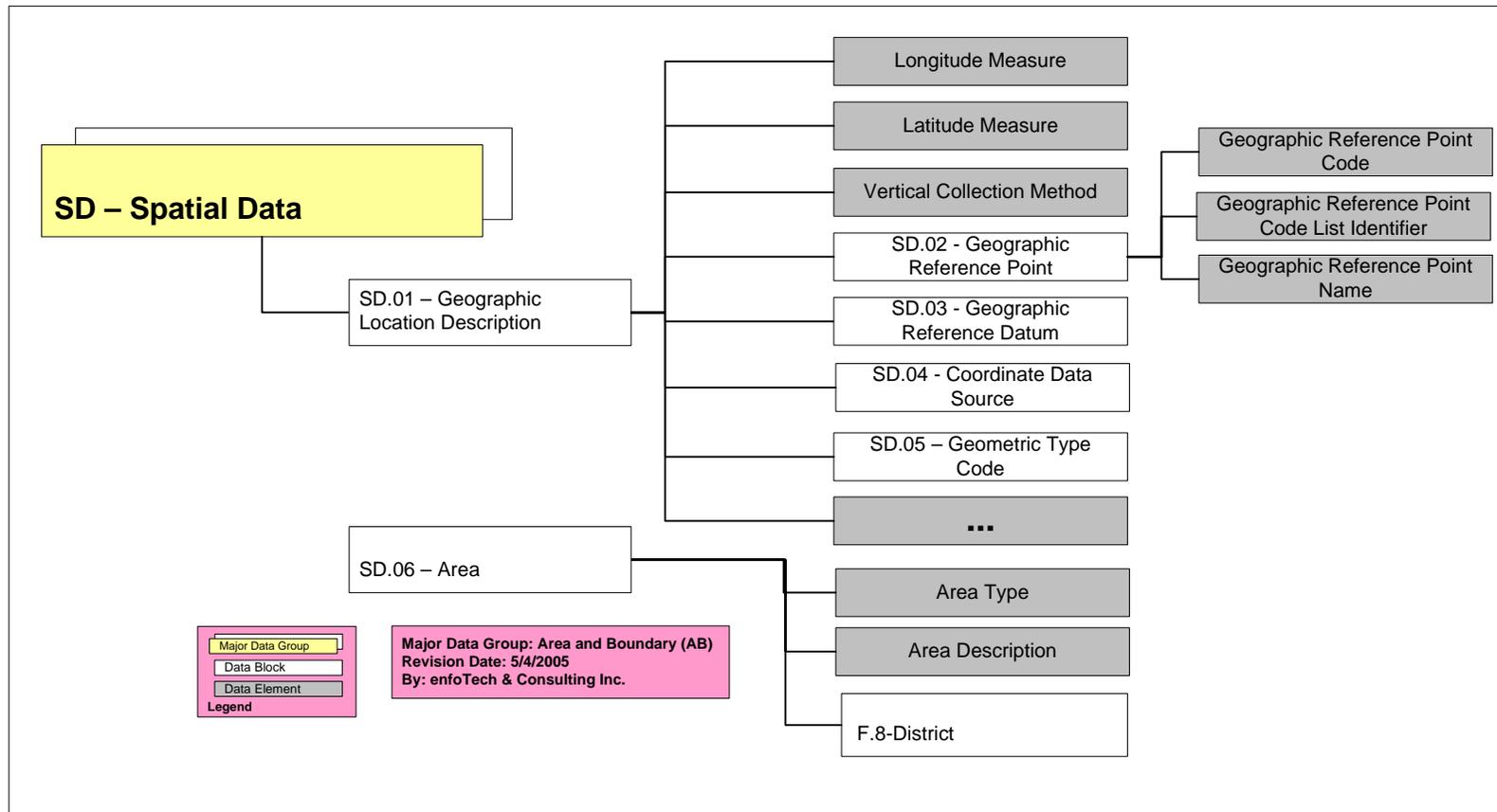


Diagram 25: Major Data Group of Spatial Data (SD)

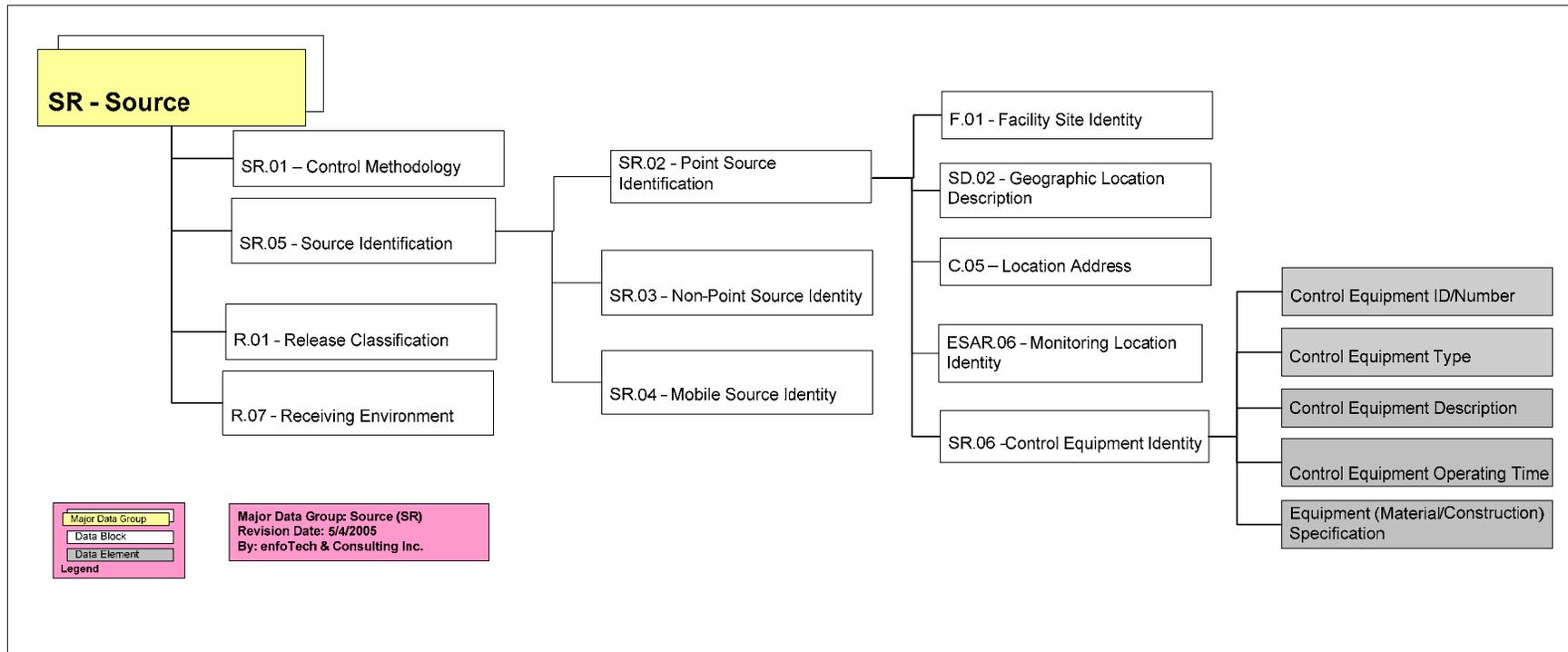


Diagram 26: Major Data Group of Source (SR)

5 CRM Inventory Analyses

5.1 Reusable Data Block Analysis

Taking inventory of Data Blocks is only the beginning of the CRM. What one does with this inventory, in terms of analysis and recognition of development opportunities really is the spirit of the CRM. The following analyses have been performed to begin the discovery and reusability process.

In order to determine the usefulness of a Data Block, a “Reusable Data Block Analysis” can be conducted to determine whether a Data Block is used, how often it is used, and where it is used among the Major Data Groups defined in the CRM. If a Data Block is created with one specific purpose for a specified data flow, the question arises as to the reusability of that Data Block. A “Reusable Data Block Analysis” can determine which data standards are lacking and which may need to be developed in the future. For instance, the Address Identification Data Block may be used many times in various Major Data Groups, but a data standard may not be available. Hence, a new data standard may be planned for development in anticipation of such a need. With this analysis in hand, one may better evaluate and prioritize the need for any future data standards development.

The following illustration shows the example Data Elements identified for the Data Block Facility Identity F.02. We can also view the Major Data Groups that this Data Block is associated with, as well as the example Data Elements that it includes.

F.02 Facility Identity		Basic identification information related to a facility site, an organization (or business entity/corporation), and/or individual person.	
Example Data Elements for F.02		Major Data Groups that use F.02	
Ownership Type		License	(L)
Facility Type		Monitoring-Ambient	(MA)
Ownership Description		Monitoring-Compliance	(MC)
Facility Status		Compliance Result	(CR)
System Classification's Unique ID/Number		Monitoring-Emergency Response	(ME)
System Classification Name		Permit	(P)
Facility Site Name		Enforcement	(E)
		Facility	(F)
		Source	(SR)
		Release	(R)
		Environmental Accident Event	(EAE)
		To Be Categorized	(TBC)
		Contact	(C)

A complete “Reusable Data Block Analysis” was conducted as part of Version 1.0 of the CRM document. Please refer to that document (still available on the Exchange Network website) for a summary of that analysis.

5.2 CRM / XML Schema Comparison Analysis

5.2.1 Facility Schema

The comparative analyses of the Facility Schema (version 2.0) with the CRM’s Facility Major Data Group reveal a close similarity in the grouping of various Data Elements. Using the current CRM’s Data Block inventory, the Facility Schema can be mapped and represented by Data Blocks. The following diagram shows how the Facility Schema may be represented by the Data Blocks:

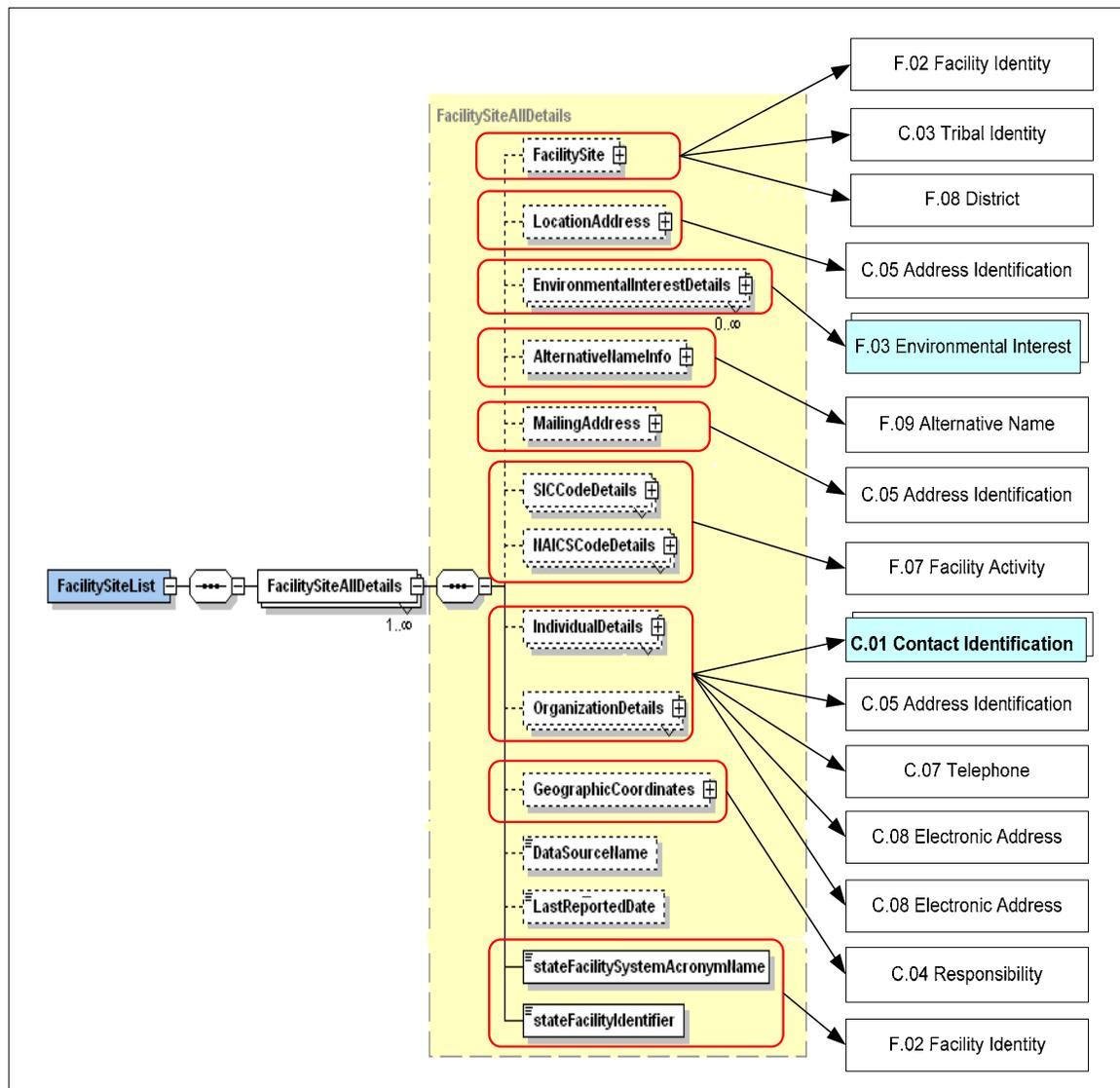


Diagram 27: Facility Schema Represented by Data Blocks

Please note the schema’s data groups shown above may contain additional Data Elements that are not part of the “example Data Elements” defined for the specified

Data Blocks. As long as the Data Elements fit into the definition of the Data Blocks, it is suitable for that Data Block.

5.2.2 e-DMR Schema (NPDES)

The electronic Discharge Monitoring Report (e-DMR) schema provides a set of core Data Elements that are common to the current DMR reporting needs of most of the state agencies. It is intended to be a national guideline for the implementation of an electronic DMR reporting system. The initial application is for DMR reporting from a permitted Facility to the State Agencies.

A complete analysis of the comparison of the Core Reference Model with the eDMR schema was conducted as part of the CRM Version 1 document. Please refer to that document for the analysis findings.

5.3 Sample Uses of the CRM to Data Standard Development

The EDSC chartered a Permitting Data Standard Action Team in late 2000 to identify and define the major areas of permitting information, and to develop a data standard that could be used for the exchange of permitting data among environmental agencies and other entities. This Action Team produced a standard that consisted of identification and tracking data believed to be universally applicable to most programs that have a permitting process or that are interested in permit related information. This standard contained the following groups of Data Elements, Permit Identification, Permitted Features, Permit Administration, and Permit Contacts. The EDSC approved that standard in December 2001.

In September 2002, the EDSC chartered a new action team (Permitting II) to broaden the existing Permit standard to include any additional information that is useful to multiple programs in order to avoid capturing similar information in more than one standard

CRM Version 1.0 conducted an analysis to demonstrate how the CRM inventory can be used to support the **Standard Data Elements for Permitting (Draft Permit II)**. For detailed information about the **Standard Data Elements for Permitting (Draft Permit II)**, please visit <http://www.epa.gov/edsc>.

6 Example Applications of the Core Reference Model

With the publication and update of the Core Reference Model, the applications for its use will become more evident. Initially, the CRM will aid in the development of the three types of environmental data flows:

- Facility-to-State Data Flow
- State-to-USEPA Data Flow
- Facility-to-USEPA Data Flow

The original CRM Workgroup conducted an analysis using an example of each of these data flows.

6.1 Facility-to-State Data Flow (Environmental Reports and Forms)

Facilities are required to submit data to their regulating state agency. The CRM can help to develop XML schema that groups the data to be submitted in a logical way, by creating Data Blocks that will be reused for many reporting data flows.

Environmental reporting data of this type is typically generated by the facility (industry, wastewater facilities, laboratories, etc.) and then submitted to the state regulatory agencies. Examples of these types of flows include discharge monitoring reporting (DMR), drinking water reporting (DWR), air emissions inventory reporting, online permit applications, and many others.

The CRM I document contains an analysis of the application of the CRM to New Jersey's Air Pollution Control Permit (Title V) Application, using CRM data blocks. The permit application form was mapped to the CRM data blocks for Facility general information and Contact information.

The analysis found that the Air Pollution Control Permit (Title V) Application form for the State of New Jersey contained information related to 32 of the CRM I data blocks.

For detailed information about this analysis, please refer to the CRM Phase I document, available at the Exchange Network website.

6.2 State-to-USEPA Data Flow

Some environmental data that is reported to the states must also be submitted to the USEPA. By identifying those CRM Data Blocks representing data submitted to the state, the task of formatting and sending data to the USEPA may possibly be reduced. As long as both agencies are aware of the definition of the data being sent, the electronic submission of data should become more efficient.

An example of State-to-USEPA data flow is the reporting of state NPDES information to EPA's Permit Compliance System (PCS) database, which was analyzed by the CRM I workgroup at the table and field level, to ensure that each piece of data within the PCS will have a home within the CRM Data Blocks.

The **Permit Compliance System (PCS)** is an USEPA database tracking information pertaining to water discharge permits issued under the National Pollutant Discharge Elimination System (NPDES). The Clean Water Act (CWA) requires that all discharges from any point source into waters of the United States must obtain an NPDES permit. Thus, the NPDES permit program regulates direct discharges from municipal and industrial wastewater treatment facilities, industrial point sources, and concentrated animal feeding operations, that discharge into wastewater collection systems or receiving waters.

States with an NPDES delegation can issue NPDES permits to regulated facilities. States with a delegated NPDES program are required to enter and update the information on NPDES activities into the PCS, which is an example of data flowing from a state to the USEPA. These activities include tracking issued permits, permit limits, monitoring data, compliance activity, enforcement activity, and other data on the regulated facilities.

The CRM I document contains an analysis of the application of the CRM to the PCS data flow, using CRM data blocks. The database tables and fields were mapped to the CRM data blocks.

The analysis found that the PCS data flow contained information related to 51 of the CRM I data blocks.

For detailed information about this analysis, please refer to the CRM Phase I document, available at the Exchange Network website.

6.3 Facility-to-USEPA Example: RCRA Permit Application

The **Resource Conservation and Recovery Act (RCRA)** requires anyone who owns or operates a facility where hazardous waste is treated, stored, or disposed to have a RCRA hazardous waste permit issued by the U.S. Environmental Protection Agency (US EPA). As an example, Part A of the RCRA hazardous waste permit application (EPA Form 8700-23, May 2002)⁹ was evaluated and mapped with the CRM Data Blocks by the CRM I workgroup.

⁹ Part A of the RCRA hazardous waste permit application can be found at <http://www.epa.gov/epaoswer/hazwaste/data/form8700/form.htm>

Based on the preliminary analysis of the RCRA Hazardous Waste Permit Application (Part A), the analysis found that the RCRA data flow contained information related to 28 of the CRM I data blocks.

For detailed information about this analysis, please refer to the CRM Phase I document, available at the Exchange Network website.

6.4 Other Uses

In addition to traditional reporting of data from facilities to States and the USEPA, there may be other uses for the CRM for non-traditional data flows; which may include cross media uses and cross program uses.

6.4.1 Multi-media Uses

Currently, the practices used to exchange air monitoring and compliance data have only a small relationship to the exchange of water compliance data. Using the current practices, results obtained from the air data flows cannot be combined easily with results obtained from the water data flows.

For example, facility data in the air permit is very similar to the facility data used in the water permit. Schemas developed for the NEI data flow might contain similar facility data elements as those for the e-DMR data flow, but the data element groupings are very different. We could expand the analysis to include the XML schema design. For example, the e-DWR (Drinking Water Report) schema includes a Lab Data Module, based on the ESAR data standard, which was created to define laboratory identification and test results. This Lab Data Schema Module is different from the one(s) developed for the NEI data flow.

The Exchange Network will work toward the objective of eliminating this deficiency and improving interoperability of data flows. Although it is recognized that there is a need to report data in different formats in order to meet specific program needs, but portions of the underlying data schema/objects contain common data elements that are similar and could be derived from a set of common Data Blocks. Common data elements and data blocks are the key resources for data sharing among multi-media programs.

The Shared Schema Components, created from the EDSC data standards, CRM, and XML Design Guidelines, are tools that can provide the foundation for Network partners from different program areas to share and assemble information to form a comprehensive environmental view for each permitted facility or for an overall regional environmental assessment.

6.4.2 One Program Area to Another Program Area

There may be situations where one program area could use information that has been obtained by another program area. In these cases, having a standard procedure to

“link” different data flows would be useful. The linkage could be accomplished by using certain key data elements within common Data Blocks that are shared by various data flows.

For example, almost all program areas use the Facility Schema Modules. If there is a need to evaluate the total potential releases (including air emissions, wastewater discharge, waste disposal, etc.) for one facility allowed under all of its existing permits, the standard Facility Schema Module using the common CRM Data Blocks and the “key data elements” could be used to combine information from different program areas.

7 Recommended Future Steps

During the development of the Core Reference Model, the CRM workgroup conducted the following steps:

- 1.) Conducted an extensive census of Data Elements from common State and EPA regulatory forms, flows, and systems.
- 2.) Compared and grouped these elements into common logical blocks, and blocks of blocks.
- 3.) Organized these blocks into a set of familiar Major Data Groups.
- 4.) Completed preliminary model CRM
- 5.) Harmonized the data model with EDSC Data standards
- 6.) Developed shared schema components (SSC) as reusable XML schema based on the data blocks that had been approved by the EDSC

The resulting preliminary model suggests the following:

- Although still preliminary, and subject to ongoing refinement, the current groups and blocks represent a sufficient critical mass of common, cross-cutting concepts to be valuable. The groups should be familiar to any practitioner who has modeled this data.
- Already some XML schema such as the TRI and OWWQX XML schema have been developed using the SSCs developed based on the CRM. As a result, many Data Blocks appear to be good candidates for common, re-usable XML schema fragments (i.e., element definitions), that can be extended and restricted as required by their program/media/project context. Continuation of shared schema components for highly reusable data blocks will be valuable.
- The CRM differs from existing and proposed EDSC standards in some areas. Rather than "force fit" the Phase I model to the standards, the group believes it better to evaluate each model as it is refined and reconstructed.

Although a significant product itself, the CRM workgroup believes that in order to make progress on the broader vision of an Exchange Network of harmonized, consistent XML schema, work on several fronts is required. The current product, while suggestive, is not yet "easy to use" for XML schema designers. Since the CRM was envisioned as a framework document, many of the required next steps are inter-dependant. Some of these follow-on efforts are best organized as a "Phase II of the CRM," others will be pursued in parallel with their respective workgroups, and some will be a mix of these. Efforts directly related to the CRM effort are:

- EDSC data standards development and review process
- XML schema guidelines development team
- XML schema review process development team

The following sections describe recommended follow-on activities and their respective leads. Please note that these activities are not represented sequentially, and may be conducted in parallel and/or in combination.

7.1 Resolve Inconsistencies Between CRM and EDSC Data Standards

Proposed lead: Joint EDSC and CRM Phase II Team

Expected Duration: 1-3 Months

Although the CRM and EDSC data standards were harmonized in late 2003, as a result of continued data standards development, further harmonization is needed, including:

- Identify significant differences in structure between current and proposed data standards and the CRM.
- For the differences identified, propose a preferred structure (i.e., revisions to the CRM, to the Standard, or both).
- For areas where the CRM proposes Data Blocks not yet covered by existing standards, a scope for new standards should be developed or recommended (especially for Data Blocks with high potential for re-use). See also the recommendation below that new XML schema fragments be developed for consensus Data Blocks. In some cases these XML schema may merit their own standards, in others they may not.

One area that needs significant attention is in the area of environmental sampling and analysis. CRM currently contains redundant Major Data Groups. The Monitoring Ambient, Monitoring Compliance, and Monitoring Emergency Response data groups were introduced in CRM Version 1.0, but since then Shared Schema Components have been created that align more closely with the EDSC Data Standard of ESAR (Environmental Sampling, Analysis, and Results). As a result, multiple Major Data Groups exist in CRM Version 2.0 that overlap in terms of business area.

7.2 Define CRM's role in the Exchange Network

Proposed Lead: New ad-hoc TRG/EDSC group

Expected Duration: 2-4 months

Momentum and activity in schema creation continues to grow, but basic questions about how some Exchange Network pieces fit together remain unanswered. Until now, data standards continue to be created and defined without the use of CRM as guidance, although the CRM Version 1.0 has been published for over 18 months. The CRM is not a part of the EDSC data standards development. In addition, CRM's role is unclear or not enforced with respect to XML schema development. Although conformance with the CRM is a review step for all Exchange Network XML schema development, compliance is not strictly enforced (as it is self-enforced) and usually occurs after schema have been created. The following areas need attention:

- Define in specific terms the role of the CRM in the Exchange Network, both in terms of:
 - Data standard selection (for future standards development)
 - Data standards development
 - XML schema development and compliance

This framework must provide guidance on how these components of the Exchange Network would, could, or should work together. It would also identify areas where we simply do not know enough to make any firm Exchange Network-wide decisions.

Over the next 1-2 years we expect much progress from work going on in broader standards efforts (such as ebXML and UML) and their related technologies. Together with our own Exchange Network experience these should allow us to revisit the provisional framework with new tools. Given that the CRM is envisioned as a major XML schema organizing tool, its success is dependant on some early agreements on the above four components.

In addition, this framework should allow for the CRM knowledge sharing and outreach, in order to convey the meaning and principals of the CRM to the Exchange Network community.

7.3 Launch CRM Phase III Development (based on item above)

Proposed Lead: CRM Phase III Team

Expected Duration: 4-6 months

As indicated above, Phase I demonstrated the feasibility of Data Block level compatibility of XML schema, but stopped short of providing the roadmap to how the CRM would get us there. Phase II demonstrated the application of the CRM for XML schema development through the creation of the Shared Schema Components.

Phase III of the CRM model would build on the efforts described above (in some cases they may be simultaneous) to build out the Model itself. The workgroup has identified candidate components of CRM Phase III; many of these components were identified in the planning effort that produced Phase I, but were held pending its completion. They are presented here in no particular order:

- Identify and model (at a high level) the relationships between Major Data Groups/Blocks, with a special emphasis on the key fields needed to establish these relationships.
- Prototype the expression of these relationships and the Data Blocks in UML (e.g., an object/class diagram)
- Develop real draft XML schema chunks for most Data Blocks as a further reality check on the utility of re-usable fragments
- Fully apply the resultant CRM, in detail, to the XML schema selected for the Schema Review Pilot. The analysis done in Phase I was mostly illustrative it did not compare these flows, element by element, with the CRM.

- Conduct a series of pro-active CRM/XML schema development workshops where XML schema developers in the early stages of XML schema development are sought out and worked with to test-drive the CRM. These workshops would serve multiple purposes: outreach, education, and validation/demonstration of the Model.
- “Lead by example” by fully implementing CRM consistent schema set for a major data flow (e.g., the Laboratory Data Standard).
- Develop a CRM document and online search tool linked to the Network Registry
- Development of a CRM Implementation Guide.
- Development of a criteria worksheet to evaluate XML schema conformance with the CRM (for uses in the TRG Schema Review Process)
- Develop an outreach and communication approach for the CRM Phase II products.

Note that in addition to the CRM implementation guide identified above, the TRG now recognizes the need for a single guide to cover all aspects of developing and utilizing XML for network flows. This would incorporate guidance currently being developed, such as the Schema Design Guidelines, Node Protocol, Schema Review Process, and additional guidance to be developed. Examples of additional guidance includes the use of digital signatures with XML documents, guidelines for developing XML style sheets for the display of XML data conveyed in Network schema, and guidance on developing and submitting reusable XML artifacts.

7.4 Responsible Parties for CRM Maintenance

The CRM is a “living” document and will require maintenance as the Exchange Network evolves. Already, the CRM has lost some relevance from its original publication in 2003. Modeling of environmental data evolves over time: the CRM will also need to evolve over time to accommodate changing business practices, technology innovations, and (most importantly), changing data standards. Without consistent updates, the work that has gone into developing the CRM will be lost.

8 Appendices

8.1 Definitions and Abbreviations

Term	Definition
CAS	Chemical Abstracts Service
Compound Data Block	A special class of Data Block that consists of data elements as well as other smaller Data Blocks.
CRM	Core Reference Model
Data Block	A grouping of related Data Elements that can be used and reused among different information flows
Data Element	A single unit of data that cannot be divided into more fundamental segments of data, and still have useful meaning
DET	Data Exchange Templates
DMR	Discharge Monitoring Reporting
DWR	Drinking Water Reporting
ECOS	Environmental Council of States
e-DMR	Electronic Discharge Monitoring Reporting
EDR	Environmental Data Registry
EDSC	Environmental Data Standards Council
e-DWR	Electronic Drinking Water Reporting
Exchange Network	Environmental Information Exchange Network
HAPs	Hazardous Air Pollutants
IMWG	Information Management Workgroup
Major Data Group	A grouping of Data Blocks and Compound Data Blocks used to represent data flow for a specified business area
NAICS	North American Industry Classification System
NEI	National Emissions Inventory
NIF	NEI Input Format
NPDES	National Pollutant Discharge Elimination System
NSB	National Steering Board
OWWQX	Office of Water Water Quality eXchange
Partners	Exchange Network Partners, such as US Environmental Protection Agency, States, and Tribes
PCS	Permit Compliance System
SIC	Standard Industrial Classifications
SoR	System of Registries
TPA	Trading Partner Agreement
TRG	Technical Resource Group
TRI	Toxics Release Inventory
US EPA	United States Environmental Protection Agency
XML	eXtensible Markup Language
XML Schema	XML Schemas express shared vocabularies and allow machines to carry out rules made by people. They provide a means for defining the structure, content and semantics of XML documents.
XML Namespace	An XML namespace is a collection of names, identified by a URI reference, which are used in XML documents as element types and attribute names. XML namespaces differ from the "namespaces" conventionally used in computing disciplines in that the XML version has internal structure and is not, mathematically speaking, a set.

8.2 References

1. 2000 Toxic Release Inventory Report for the State of Michigan, (see <http://www.epa.gov/tri/tridata/tri00/state/Michigan.pdf>)
2. Air Pollution Control Permit (Title V) Application of New Jersey, (see <http://www.state.nj.us/dep/aqpp/applying.html>)
3. Environmental Data Registry (EDR), (see www.epa.gov/edr)
4. Environmental Data Standards Council (EDSC), (see www.epa.gov/edsc)
5. Environmental Information Exchange Network, (see www.exchangenetwork.net/common/default.asp)
6. National Emissions Inventory Input Format (NIF), (see www.epa.gov/ttn/chief/nif/index.html#ver2)
7. Permit Compliance System (PCS), (see www.epa.gov/ceisweb1/ceishome/ceisdocs/pcs/pcs-exec.htm)
8. RCRA Hazardous Waste Part A Permit Application (EPA Form 8700-23), (see www.epa.gov/epaoswer/hazwaste/data/form8700/forms.htm)
9. Toxics Release Inventory (TRI), (www.epa.gov/ceisweb1/ceishome/ceisdocs/tri/tri-exec.htm)
10. Central Data Exchange (CDX), (see www.epa.gov/cdx)
11. Water Discharge Permits (PCS) tables in Oracle Database format used by Envirofacts, (see www.epa.gov/enviro/html/pcs/pcs_table.html)
12. XML Registry, (see [http://oaspub.epa.gov/emg/xmlsearch\\$.startup](http://oaspub.epa.gov/emg/xmlsearch$.startup))
13. Facility Registry System Schema, (see http://oaspub.epa.gov/emg/portal.navigate?P_LIST_OPTION_CD=XMLCSALL&P_REG_AUTH_IDENTIFIER=1&P_DATA_IDENTIFIER=89551&P_VERSION=1)
14. Discharge Monitoring Report (e-DMR) Schema, (see http://oaspub.epa.gov/emg/portal.navigate?P_LIST_OPTION_CD=XMLCSALL&P_REG_AUTH_IDENTIFIER=1&P_DATA_IDENTIFIER=89553&P_VERSION=1)
15. Environmental Council of States (ECOS), (see www.sso.org/ecos)
16. Nebraska Department of Environmental Quality's Permitting Application, Inspection Form, Authorization Form, and etc. provided by Dennis Burling of Nebraska DEQ, (see www.ndeq.state.ne.us/)
17. Delaware Department of Natural Resources' Permitting Application, Inspection Form, Authorization Form, and etc. provided by Dennis Murphy of Delaware DNR, (see www.dnrec.state.de.us/dnrec2000/)
18. Michigan Department of Environmental Quality's Permitting Application, Inspection Form, Authorization Form, and etc. provided by Michael Beaulac of Michigan DIT, (see www.michigan.gov/deq/)