**Facility Data Integration Discovery and Analysis**

Version 1.00

April 20, 2016

Table of Contents

Table of Contents i

List of Figures ii

List of Tables ii

Revision Log iii

Acknowledgements iv

1 Introduction 1

1.1 E-Enterprise for the Environment 1

1.2 Approach 3

2 Current State 4

2.1 Problem Statement and Use Cases 4

2.2 Use Cases 6

2.3 IPT Member Survey Results 6

3 Discovery Session Findings 14

3.1 Approach and Information Types Description 14

3.2 Missouri Discovery Session 14

3.3 Wyoming Discovery Session 21

3.4 New Jersey Discovery Session 25

3.5 EPA/FRS Current State and Future Plans 30

3.6 Summary 33

4 Facility Data Management Partnership Vision 37

4.1 Users 37

4.2 Principles 37

4.3 Components 41

4.3.1 API First 42

4.3.2 Common Enough Information Model 42

4.3.3 Some Shared Business Rules 42

4.3.4 Sharing Good Practices and Tools 42

4.4 E-Enterprise Facility Governance 42

5 Summary and Recommendations for Phase II 44

5.1 Summary of Findings 44

5.2 Phase II Recommended Approach 44

5.3 Phase II Recommended Actions and Goals 44

6 References 46

Appendix A. Charter for an Integrated Project Team to Scope Options for Integrating Facility Identification Data 47

Appendix B. Use Cases 52

Appendix C. Discovery Session Introductory Letter 54

Appendix D. Discovery Session Agenda 56

Appendix E. Burden Reduction for Direct Reporters – the CDX/FRS Facility Management Widget and TRI 59

List of Figures

[Figure 2‑1 Level of Maturity in the Evolution of Integrated Facility Data Management 7](#_Toc448850879)

[Figure 2‑2 Data Reconciliation Timing 7](#_Toc448850880)

[Figure 2‑3 Data Presentation for Selection and Correction 8](#_Toc448850881)

[Figure 2‑4 Staff Facilitation of Integrated Facility Management 8](#_Toc448850882)

[Figure 2‑5 Formal Data Governance – Policy or Procedure 9](#_Toc448850883)

[Figure 2‑6 Core Facility or Master Data Terminology 10](#_Toc448850884)

[Figure 2‑7 FRS Utilization 11](#_Toc448850885)

[Figure 3‑1 MoGEM Future Data Flow 17](#_Toc448850886)

[Figure 3‑2 MoGEM Domain Model 20](#_Toc448850887)

[Figure 3‑3 NJDEP Architecture 26](#_Toc448850888)

[Figure 3‑4 NJEMS Architecture 26](#_Toc448850889)

[Figure 3‑5 NJEMS Logical Data Model 29](#_Toc448850890)

[Figure 3‑6 Known FRS Issues and Proposed Approaches 32](#_Toc448850891)

[Figure 3‑7 FRS ‘Thing of Interest’ Model 33](#_Toc448850892)

[Figure 4‑1 Facility Data Sharing Principles Outline 38](#_Toc448850893)

[Figure 4‑2 IPT Response to Straw Facility Data Sharing Principles 39](#_Toc448850894)

List of Tables

[Table 2‑1 Lessons Learned 12](#_Toc448850900)

[Table 2‑2 Challenges for Collaboration on Management of Facility Data 13](#_Toc448850901)

[Table 3‑1 Missouri Discovery Session Business Process Information 18](#_Toc448850902)

[Table 3‑2 Wyoming Discovery Session Business Process Information 23](#_Toc448850903)

[Table 3‑3 New Jersey Discovery Session Business Process Information 27](#_Toc448850904)

Revision Log

| **Date** | **Version No.** | **Description** | **Author** |
| --- | --- | --- | --- |
| March 14, 2016 | 0.01 | Delivery of Facility Data Integration Discovery and Analysis Rough Draft | Luke Gentry, CGI Federal Lydia Lyshevski, CGI Federal Louis Sweeny, Ross Strategic |
| April 20, 2016 | 1.0 | Delivery of Facility Data Integration Discovery and Analysis – Addresses Comments from IPT Team, Collaborative Discussion, and Co-Chairs | Luke Gentry, CGI Federal Lydia Lyshevski, CGI Federal Louis Sweeny, Ross Strategic |
|  |  |  |  |
|  |  |  |  |

Acknowledgements

The Facility Data Integration Integrated Project Team (IPT) developed this Discovery and Analysis with the support and contributions of a number of individuals and programs within EPA and state agencies. The IPT acknowledges:

| **State and EPA Facility Data Integration IPT** | |
| --- | --- |
| Kimberly Hoke, MO DNR, Co-Chair  Wayne Begin, Kentucky DEP  Jim Bridgewater, New Jersey DEP  Eric Brown, Colorado DPHE  Dennis Burling, Nebraska DEQ  Joe Carioti, EPA OSWER  Scott Christian, EPA OSWER  Chris Clark, EPA OEI  Regina Crolley, South Carolina DHEC  Leslie Cronkhite, EPA OCFO  Sally Dombrowski, OAR, OAQPS  Linde Dytrich, Texas CEQ  Ron Evans, EPA OAR  Angela Feldman, EPA  Pam Galli, Rhode Island DEM  Kristen Gastner, EPA OSWER  Alex Geller, Vermont DEC  Malcom Gentry, New Hampshire DOIT  Steven Girt, Wyoming DEQ  Brian Gregory, Georgia DNR  Jen Gumert, Pennsylvania DEP  Charles Homer, Montana DEQ  Michael Hurley, Massachusetts DEP | Lee Kyle, EPA OEI, Co-Chair  Martin Husk, EPA OAR  Matthew Kelly, EPA OEI  Mike Kvitrud, Wisconsin DNR  Linda Lazich, Ohio, OEPA, Air Services  Jose Lopez, EPA Region 8, ENF-PJ  Mike Matsko, New Jersey DEP  Christopher Mentzer, Maryland DE  Annie Morales, Texas CEQ  David Mummert, Maryland DE  Theresa Pagodin, New Jersey DEP  Siobhan Perricone, Vermont ANR  Ron Ryan, EPA OAR  Tobias Schroeder, EPA  Peter Soyka, E-Enterprise Coordinator  Nadia Spice, EPA Office of Water  Gayle Stewart, Texas CEQ  Dalroy Ward, EPA OEI  Herb Ward, Virginia DEQ  Doug Waldron, Georgia DNR  Ben Way, Wyoming DEQ  Elisa Willard, Colorado DPHE |
| **Other Contributors** | |
| Michael Alford, SRA  Lynn Capuano, Exchange Network Coordinator  Alison Ferner, CGI Federal  Lucas Gentry, CGI Federal  Beth Graves, ECOS  William LaBar, CGI Federal  Lydia Lyshevski, CGI Federal | Jimmy Mahady, Ross Strategic  Greg McNelly, ECOS  Guy Outred, Windsor Solutions  Eric Snell, Linko Technologies  Louis Sweeny, Ross Strategic  Rob Willis, Ross Strategic |

# Introduction

## E-Enterprise for the Environment

E-Enterprise for the Environment is an initiative that will transform the business and delivery of environmental protection in the United States. Through the collaboration of the states, EPA, and tribes, E-Enterprise will improve environmental performance and enhance services to the regulated community, environmental agencies, and the public. The primary purpose of E-Enterprise is to improve environmental protection.

EPA and the states have worked together through the State-EPA E-Enterprise Working Group to create the vision for E-Enterprise. The first step in defining E-Enterprise was the creation of a Conceptual Blueprint that defines the design and operating principles, as well as six primary E-Enterprise Components.

### E-Enterprise Design and Operating Principles

The ten principles identified in the Blueprint are a combination of assertions about the operation of E-Enterprise and assumptions regarding system capabilities.

1. Partnership of Environmental Government Regulators
2. Honoring Delegated Authority
3. Modernize and Improve Environmental Regulations and Programs, and their Implementation
4. Joint Governance Prioritizing Activities
5. Emphasizing User Perspective
6. Creating and Expanding Systems to Improve Two-way Business Transactions
7. Interoperability of Partner Systems and Partner Use of EPA Systems
8. Open Data and Web Services
9. Advanced Monitoring Technologies and New Data Collection and Analysis Techniques
10. Shared Technical and Programmatic Infrastructure

See Section 3 of the Blueprint for additional content and context regarding these principles.

### E-Enterprise Components

The Blueprint identified six primary components that will make up E-Enterprise.

1. Modernizing and Streamlining Programs and Regulations
2. Portfolio of Advanced Monitoring Technologies
3. The E-Enterprise Portal
4. Partner Access and Transaction Systems
5. Open Data and Web Services
6. E-Enterprise Shared Technical and Programmatic Infrastructure

See Section 4 of the Blueprint for additional content and context regarding these components.

### Governance and Coordination

The E-Enterprise Leadership Council (EELC) was formed to provide coordination and oversight for E-Enterprise activities. In cooperation with the EELC, the E-Enterprise and Exchange Network Management Board (MB) and E-Enterprise and Exchange Network Interoperability and Operations Team (IOT) supports information technology projects of E-Enterprise. The E-Enterprise Facility Data Integration IPT is chartered by the MB to provide stakeholder input into the design and development of E-Enterprise facility data integration services.

#### Facility IPT Charter

EPA, states, and tribes are working to implement E-Enterprise for the Environment—a transformative initiative that will improve the business of environmental protection in the United States. Sharing integrated facility identity information between states and EPA is critical to achieving that vision. EPA and each state environmental agency—as well as each program office (such as air, water, and waste) within those agencies— separately collect, record, and maintain both the specific data needed for each set of regulations and the core information to identify each regulated facility by name, address, geographic location, owner, etc.

As a result, facilities have to report and update the same basic identification data to multiple programs, at multiple levels of government, and at multiple times. Oftentimes, there are differences and discrepancies among these disparate sets of facility identification data (synonymous with facility data management) that prevent accurate correlation across programs and agencies.

Environmental agencies have been independently working to better integrate their facility information. EPA uses the Facility Registry Service (FRS) to gather, link, and correct facility data from individual program systems at EPA. Many state agencies have invested in their own similar systems for integrating facility identification information at the state level.

Despite improvements in methods to share, link, and correct facility information through back-end reconciliation, data on regulated facilities remains highly fragmented and incomplete. E-Enterprise offers an opportunity to explore possible approaches to integrating facility identification data across programs and agencies and correcting facility data as it is being reported. A more comprehensive approach or approaches can:

* Save time, staff, and money for both agency regulators and regulated facilities by allowing them to streamline data collection and reporting requirements and to assemble more quickly the multi-media environmental data they need for consolidated reports, permits, and inspections.
* Provide the public a more complete understanding of all the regulatory obligations and environmental impacts at each facility across media programs.
* Help agencies to manage their responsibilities more efficiently by identifying the most serious risks, setting priorities, and establishing performance measures.
* Increase data accuracy and thereby reduce the risk of disseminating incorrect information about a regulated entity.

The full charter is available for review in Appendix A. Note that terminology has been refined throughout the process – ‘facility identification data’ has become synonymous with ‘facility data management.’

## Approach

Using the E-Enterprise Architectural Principles as guidance, the Facility Data Integration IPT brought together E-Enterprise partners from EPA and states. The IPT solicited and documented information about facility data integration focusing on identifying collective use cases, requirements, and overall functionality documented in this Discovery and Analysis document. These materials will serve to inform the next steps in the development of facility data sharing solutions.

### IPT Activities

In order to achieve the primary objective of gathering input and requirements from EPA, state, and local government stakeholders to find ways to share facility profile data, the group engaged in the following activities:

* Gathered information about a diverse set of partner experiences related to facility master data management (MDM).
  + Catalogued existing partner systems and summarized their features and characteristics.
  + Collected information about lessons learned on system implementation, business process changes, stewardship, and policy governance.
* Conducted multiple facilitated discussions on problem statements, opportunities, key hypotheses, challenges, and constraints.
* Conducted detailed one-on-one discovery sessions with three state members to allow for a deeper dive into business processes, data models, technical approaches, lessons learned, and challenges.
* Began to develop use cases for facility data integration and data sharing.
* Began to compare state business rules with existing EPA FRS services and other EPA program business rules.
* Authored recommendations for next steps and possible pilot options that can be used to inform follow-on phases of the IPT (this report).

### Scope of Document

The purpose of this document is to capture the information gathered by the IPT and to provide a foundation upon which subsequent phases of work can be built. Possible activities for subsequent phases are identified in Section 5 of this document.

# Current State

This section provides details on the current state of facility data management approaches of IPT members. The information presented here was gathered through IPT meetings as well as detailed discovery sessions held with three IPT members representing three different states of maturity related to facility master data management.

## Problem Statement and Use Cases

Today, data is one of the most valuable assets in the business of environmental protection, human health, and ecological stewardship. To respond to the growing demands of US citizens, global environmental management, and the regulated community, agencies must have access to tremendous amounts of accurate data as well as the means to effectively leverage collected information. Consequently, data integration is among the most significant issues that is currently being addressed. Effective facility data integration strategies and tools will support efficient use of resources, insurance of data accuracy, and utilization of vast data stores to help achieve strategic goals (ex. compliance assurance, historic trending, and safety monitoring).

Today, EPA’s FRS uses sophisticated algorithms to integrate and correct facility data gathered across EPA. Facility and related environmental interest data, that feeds into FRS, flows from disparate data sources into the agency via multiple program office entry points including EPA direct reporters as well as Exchange Network (EN) data flows from trading partners such as state, local government, and tribal (SLT) environmental agencies and departments. This data includes information on facilities, locations, environmental interests, contacts, corporate owners, and industry classifications.

SLTs and other trading partners utilize different methods, best practices and commercial off the shelf (COTS) software products to integrate local facility information across regulatory programs within their organizations. States and EPA have over 20 years of experience with taking various approaches to facility data management. State agencies have encountered a wide range of outcomes, including successes, ongoing challenges and abandoned efforts. These disparate approaches and varying levels of program maturity related to master data management present a unique challenge to the vision and objectives associated with E-Enterprise for the Environment. Early work was focused on establishing automated mechanisms for batch sharing of facility data via the FacID EN flow. While many states implemented this flow, the next level of business integration with EPA was never completed. State facility integration represented in FacID is reported to EPA. And while states can now access FRS services using their state IDs, these state IDs are not used in FRS integrations. Finding ways to use the same facility IDs and profile information is one of the key goals of this IPT. Additionally, facility data from these trading partners is reported with long delays after its initial collection, with varying frequency and levels of completeness. This can lead to data becoming dated or possibly even incorrect. For the first time, detailed information on state facility interests and a largely comprehensive inventory of state facility interest types were explored by the IPT.

Recently, EPA has implemented real time data integration between FRS and the Central Data Exchange (CDX) Web reporting interface, for direct reporters to EPA, for efforts such as Office of Air and Radiation’s (OAR) Compliance and Emissions Data Reporting Interface (CEDRI), Office of Chemical Safety and Pollution Prevention’s (OCSPP) Toxic Substances Control Act (TSCA) related data flows, and EPA’s Toxics Release Inventory data collection. These CDX/FRS shared services allow for any program office data collection across the Agency to enable direct reporters to look up, associate, correct, and add new facility data directly to program office systems and FRS simultaneously. This type of integration creates the ability to update or correct facility data as it is being reported and is significantly lowering regulatory burden, increasing overall data quality, and enabling program offices across EPA, other federal agencies, and trading partners to more effectively carry out their mission. Establishing this type of connectivity between EPA and state systems will be explored in Phase II.

### Problem Statement

The *E-Enterprise Conceptual Blueprint* and IPT charter acknowledge the complexity of the challenges that this IPT seeks to address. The IPT identified numerous challenges to E-Enterprise goals for co-regulators and stakeholders related to facility data.

Today, at the regulated entity service level:

* Facilities must re-enter the same information multiple times.
* Facilities must manage and correct inconsistent linkages/accounts, or lack of linkages, between accounts across programs and different levels of government. While data may be maintained by DBAs or regulated users, in general, facility data is difficult to maintain across different systems.
* Facilities cannot access a consistent and integrated view of all their holdings, sub-facilities, accounts, users, etc.

Similarly, state and EPA (co-regulator) programs face serious challenges with management, performance, burden, and barriers stemming from the present-day management of facility data:

* Many agencies lack the information needed to better manage facilities in an integrated and high-resolution fashion for all types of business processes such as permitting, compliance and enforcement, among others.
* Agency strategy/decision-making is hampered by the fractured view of facilities. While some state system integrations include a variety of regulatory systems that have common regulated interests/facilities, a significant amount of data is fractured within the current universe.
* The state universe is numerically bigger and different from EPA. Some business problems are the same, while others are unique to local levels.
* State and EPA program data managers need systems to manage their facility data. These systems must meet unique program and state and federal regulatory needs and requirements while remaining harmonized with state and EPA agency views.
* Without a unified view of a facility and where it is, it is difficult to perform geospatial analysis, trend analysis, and planning, all of which are becoming increasingly important.

As a result of these issues, the public inherits a fractured view of facilities. This results in a lack of transparency on environmental status, activities, and performance.

The IPT recognizes that, across the nation, there are pioneers in addressing the data management problem within their enterprise. One of the fundamental goals of this IPT, inclusive of all phases, is to determine how facility data integration can be addressed across the environmental enterprise.

Through better integration/collaboration of systems, services, and business processes, E-Enterprise partners can improve facility data management and thereby address the long-standing problems summarized above. A collaborative solution will, in many cases, be superior in meeting the needs of shared stakeholders to a solution that either EPA or any one state could formulate on its own. This IPT is dedicated to evaluating ideas geared to improve facility data integration by devising practical, actionable, long-term solutions to existing problems.

The next generation of facility data management calls for:

* Coordination across programs and levels of government.
* Management of information at a sufficient level of detail (ex. several levels below facility such as features including monitoring points, smoke stacks, and plots of land) where this lower level of detail is required by regulatory requirements.

## Use Cases

To illustrate existing problems via realistic scenarios, the IPT began to identify use cases from a variety of perspectives. Regulated entity and co-regulator (ex. federal, state, tribe, local government) views were analyzed.

Use cases were not fully explored in this phase of the IPT. The initial use cases discussed are listed in Appendix B and will be candidates for further exploration in subsequent IPT phases.

## IPT Member Survey Results

In an effort to learn from all IPT participants, a member survey was conducted early in the IPT. Results were used to:

* Gauge the current state of facility data integration among participating agencies.
* Understand existing data management processes among participating agencies.
* Elicit conversation on topics of interest, concerns, and lessons learned.

As results are discussed, it is critical to take into account that information shared is self-reported.

Sixteen agencies responded to the survey. Figure 2‑1 illustrates the distribution of experience, where 4 agencies have 0 to 4 years of experience, 7 agencies have 5 to15 years of experience, and 5 agencies have over 15 years of experience. This range in familiarity with data integration is reflected in the distribution in survey response perspectives.

When asked to estimate the current status of their agency on the spectrum of evolution of integrated data facility management, the results were again highly distributed, as shown in Figure 2‑1. While 38% of the participating agencies identified themselves as mature, 38% identified themselves as intermediate and, 31% indicated they are getting started on the process of developing facility data integration processes and tools within their system(s). At every maturity level, there are unique challenges faced and lessons learned.

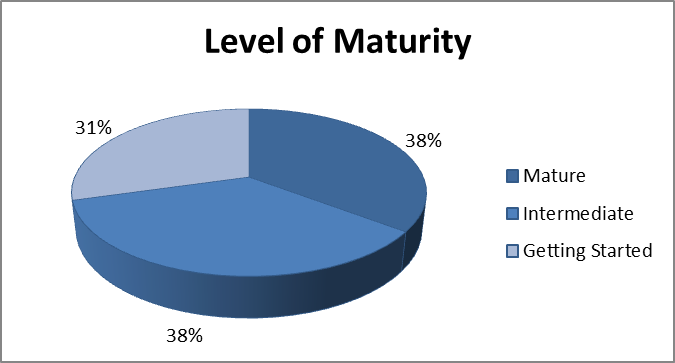


Figure ‑ Level of Maturity in the Evolution of Integrated Facility Data Management

Due to the dynamic nature of facility data, existing data will inevitably require maintenance and corrections. Agencies reconcile data at different points in the process – 20% on the front end (before records are added or updated), 40% after the fact, and 40% use a hybrid approach; refer to Figure 2‑2.

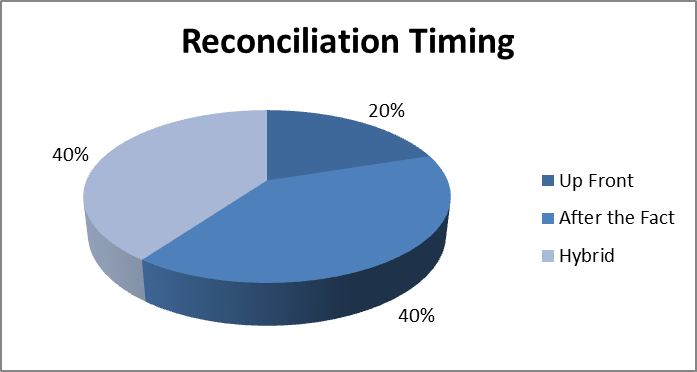


Figure ‑ Data Reconciliation Timing

Utilization and maintenance of existing data is yet another principal component of facility data integration. Figure 2‑3 demonstrates that 33% of the participating agencies present data for selection and correction to their users, 27% allow users to select data but not to update it, and 40% do not offer integrated facility data to users at all. Boosting the use and reuse of existing, vetted information is one of the goals of this IPT.

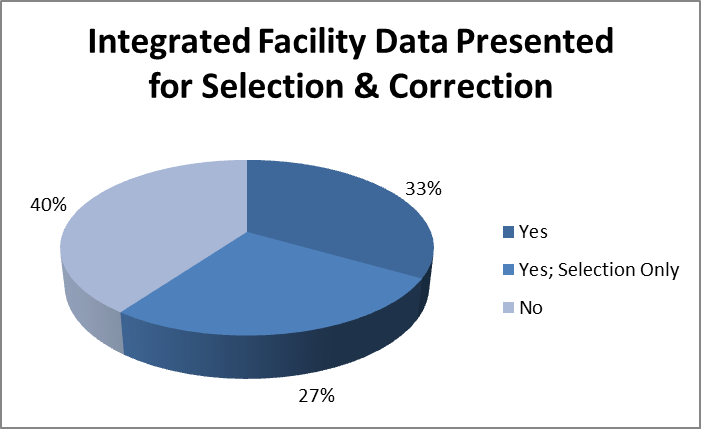


Figure ‑ Data Presentation for Selection and Correction

Of agencies with existing systems, 80% have staff facilitating integrated facility management. As shown in Figure 2‑4, 27% of these organizations have dedicated staff. Commonly, program subject matter experts (SMEs) have some form of data integration stewardship responsibilities.

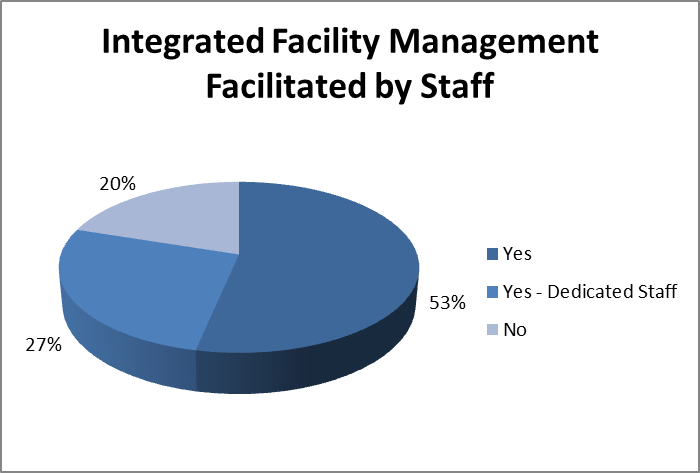


Figure ‑ Staff Facilitation of Integrated Facility Management

Data governance helps ensure consistent and reliable data via clear documentation of standards. Solid governance can also help align goals, reconcile data, define criteria/metrics, guide analytics, assure adherence to relevant policy and security guidelines, and ultimately save time and money by minimizing waste as shown in Figure 2‑5. Formal data governance policies or procedures exist at 63% of the agencies surveyed.

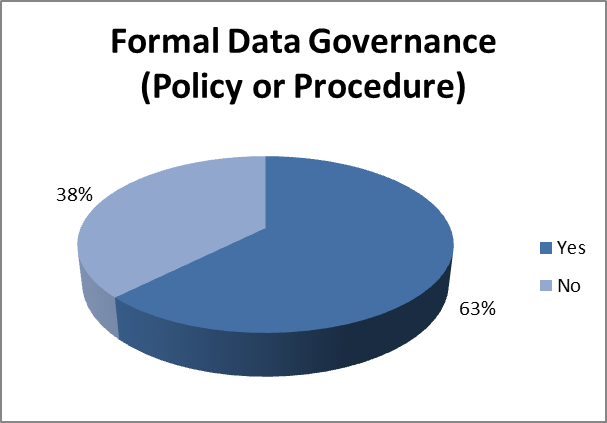


Figure ‑ Formal Data Governance – Policy or Procedure

When referring to core facility or master data, it is essential to reach agreement on and formalize terminology. As facility data integration expands across programs and states, these definitions will help avoid ambiguity and errors. The full spectrum of lingo related to master data is shown in Figure 2‑6. The most commonly used terms are facility, site, regulated entity, and environmental interest. In all, 15 nomenclatures are used.

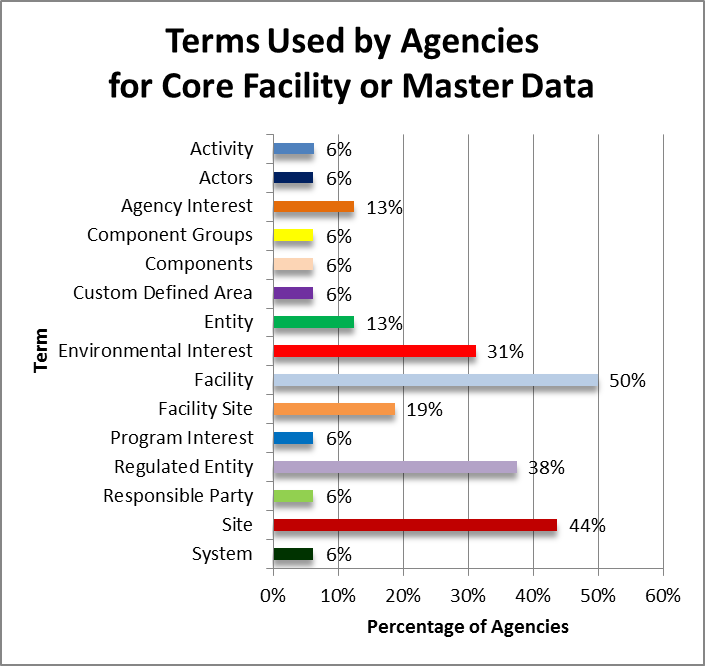


Figure ‑ Core Facility or Master Data Terminology

Agencies define master data at a variety of levels. The vast majority defines this data at the facility level; over a quarter of the participants define master data at the emission point for permitted air facilities. It is not uncommon for an organization to define data at a variety of levels to accommodate the numerous scenarios handled by the system (such as different types of permits).

Data stored in EPA’s FRS system can be used in tandem with the information collected, stored, and presented at the local level. However, as shown in Figure 2‑7, less than a quarter of the participating agencies utilize FRS today. As environmental systems evolve, the aim is to increase this type of interconnectivity among agencies at all levels of government to help assure that consistent, reliable, and up-to-date facility profile information is utilized across the nation.

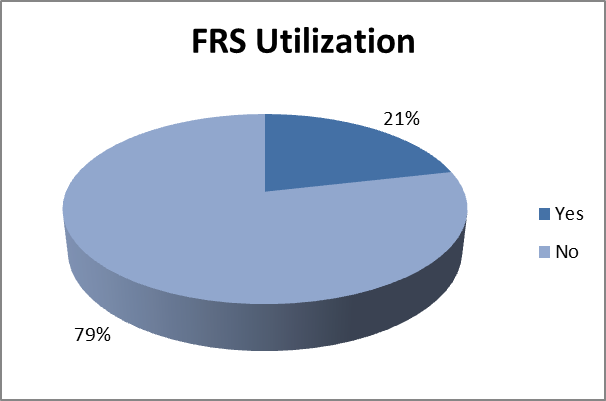


Figure ‑ FRS Utilization

The wide range of experiences among survey participants revealed valuable lessons learned as shown in Table 2‑1, where the most prevalent feedback is at the top of the table.

Table ‑ Lessons Learned

|  |  |
| --- | --- |
| **Number of Responses** | **Lesson Learned** |
| 4 | Ensure engagement among stewards/administrators to work toward a common goal and strategy – collaboration and coordination are critical |
| 3 | Buy-in from higher-up management is key |
| 3 | Facility integration is an on-going and long-term process; expect continuous changes |
| 3 | Data quality (ex. minimal duplication, data sharing/integration between agencies) must be a priority |
| 2 | Judicious decisions must be made about data that is shared vs. specialized data that is maintained within an agency/division |
| 2 | Handling data integration, data deduplication, data quality, governance, and validation tools/measures, and maintenance require a lot of effort and discipline |
| 2 | Set up user privileges, ownership hierarchies, procedures, and linkages to manage facility data in advance, to avoid spur of the moment ad-hoc changes |
| 2 | Clearly and methodically define language and terminology (ex. 'facility' or 'site') to accommodate participating programs; regulations may impact both |
| 2 | Data clean-up requires the participation of experts with hands-on knowledge of the program; input from a variety of viewpoints is important |
| 2 | Data is dynamic – plan on data having to be edited and maintained |
| 1 | Regulatory constraints on data must be considered |
| 1 | Establish a clear agreement policy between agencies/divisions pertaining to data changes |
| 1 | Set aside resources (ex. people, money) ahead of time |
| 1 | Balance between data consistency and unique needs of users is expected (i.e. original/not normalized data may need to be retained by agencies due to legislative constraints) |
| 1 | Willingness to accept change (semantic, procedural, etc.) is critical but often lacking |
| 1 | Details such as facility status, measurement location of latitude/longitude, etc. must be clearly defined |
| 1 | Data management occurs at different levels (ex. site/facility vs. company/owner) depending on the item of interest (ex. new/updated permit) |
| 1 | Tight controls and privileges are required when a wider range of users are permitted to edit master data |
| 1 | Strict rules that require integration may be necessary to promote progress |

The biggest challenges for collaboration on management of facility data were identified and enumerated, as shown in Table 2‑2. There is overwhelming agreement that facility data integration presents significant challenges at legislative, political, and technological levels. However, compelling goals, strong partnerships, and flexible solutions are expected to offset these complications.

Table ‑ Challenges for Collaboration on Management of Facility Data

|  |
| --- |
| **Topic** |
| Reconciling or understanding the differing purposes and interest in data (regulatory, scientific, public use) |
| Providing flexibility for unique circumstances |
| Regulatory differences in how facilities are defined |
| Training staff |
| Aligning information technology with established business rules |
| Determining and agreeing on the authoritative source |
| Determining how information should be maintained, shared, and used |
| Eliminating duplicate data |
| Defining and agreeing on terminology cross-state and agency |
| Getting all parties to look at the big picture |
| Making participation easy |
| Fear of losing control of data and implications at the program level |

Survey responses helped level-set expectations by clarifying the current state of data integration among IPT participants. Feedback was used to drive Phase I conversations and will help orient the IPT moving forward.

# Discovery Session Findings

During the course of the IPT efforts, detailed one-on-one discovery sessions (informally referred to as deep dives) were conducted with three state members. These sessions facilitate a more thorough examination of business processes, data models, technical approaches, lessons learned, and challenges faced by environmental organizations participating in the IPT.

## Approach and Information Types Description

IPT meetings and work group member surveys were used to gain a broad view of the status of environmental facility management efforts and approaches. In order to get a deeper, more detailed view, these one-on-one discovery sessions were conducted with three member states. Participants were selected from a pool of member states that expressed willingness to contribute to the IPT in this fashion through survey results and IPT group discussion. To constrain the scope of the efforts, three states that appeared to be at different levels of maturity, related to facility management, were invited to participate. Discovery sessions were conducted with:

* Missouri – Getting Started
* Wyoming – Intermediate
* New Jersey – Mature

In addition, Texas provided detailed information about their system by completing the discovery session agenda questionnaire. This information makes Texas a strong candidate for further analysis and possibly additional discovery session meetings in a subsequent phase of the Facility IPT.

Prior to discovery session meetings, each participant was provided an introduction letter and an agenda; the documents are provided in Appendix C and D. The letter imparted the importance of the discovery sessions and requested that the participants provide all available materials ahead of the session, if possible.

The agenda included specific questions that would be raised in the sessions, categorized by topic. In preparation, participants were invited to contemplate each topic ahead of the session.

Some materials provided by discovery session participants cannot be shared due to the confidential nature of the content. Materials that can be shared and distributed are available on the Facility IPT SharePoint site.

Information presented for each discovery section is accurate as of the time of the respective discovery session meetings.

## Missouri Discovery Session

The Missouri discovery session was held by conference call on November 10, 2015. The following materials were provided by Missouri either prior to or shortly after the discovery session.

| **Missouri Discovery Session Materials Provided** |
| --- |
| MoGEM Conceptual Solution Architecture, Technology Roadmap and Architecture Plan  MoGEM Current Status – November 2015  MoGEM Governance Board Charter  MoGEM Governance Operation Plan 1.3 – FINAL  MoGEM Future Vision v1.2 – FINAL  MoGEM Logical Data Model v1.1 FINAL  MoGEM Technology Roadmap and Architecture Plan v1.2 – FINAL APPROVED  Roadmap for printing – Final approved |

### Participants

| **Missouri Discovery Session Participants** | |
| --- | --- |
| Kimberly Hoke, MO DNR  Mark Finnell, MO OA ITSD | Lee Kyle, EPA OEI  Lucas Gentry, CGI Federal |

### Level of Experience/History

Missouri is in the planning stages of establishing master regulated entity data. Today, Missouri Department of Natural Resources’ (DNR) regulated entity data is stored in numerous information silos. DNR’s *Future Vision* document describes their current situation as:

DNR currently uses separate applications across different program areas to capture master data for a regulated entity. These program applications are built on platforms such as DB2, SQL Server, Access, and Excel. Most of these applications do not have the capability to transfer master data electronically from one program application to another program application. For example, if there is a name change for a regulated entity in one program application, this update does not automatically occur for the same regulated entity captured in another program application. Since the master data for the same regulated entity is being captured in different applications across different program areas, it has become difficult for DNR to accurately report on the number of regulated entities and specific activities occurring at a single regulated entity. In addition, the master data for these regulated entities captured in the applications are not always accurate. The need has been identified to improve the data management of the regulated entity subject area to provide clean, consistent master data to the enterprise.

Missouri discovery session participants indicated that tackling the ‘facility problem’ is foundational to addressing DNR reporting requirements as well as to the need to maintain and modernize legacy applications.

Missouri was aware that master data management approaches were in place and proven effective in other domain areas (such as finance), investigated approach options, and is now pursuing a master data management approach for their environmental regulated entity data. The effort is known as the Missouri Gateway for Environmental Management (MoGEM).

### Technical Approaches

Missouri is currently in the process of a comprehensive effort to define and implement their master data management approach for MoGEM. A detailed architecture and roadmap are provided in the materials that their project is currently generating. The information presented here is a summary based on MoGEM materials and Missouri discovery session discussion.

The *MoGEM Technology Roadmap and Architecture Plan* summarizes DNR’s regulated entity data management architecture approach as:

The MDM architecture type that was adopted during the analysis phase was a combination of ‘Consolidated’ and ‘Collaborative’ meaning that we will store a ‘single version of truth’ master data record in a central repository while also having each program application store the version that was created within their system.

Figure 3‑1, and associated description, depicts the future data flow envisioned for MoGEM. Additional information can be found in DNR’s *Future Vision* document.

The figure shows an example of how data will flow for a Regulated Entity (Doe Run Herculaneum) which is regulated by Air Pollution Control Program, Water Protection, and Hazardous Waste Program. The Doe Run Herculaneum record exists in MoEIS (Air Program), MoCWIS (Water Protection), and Fees & Taxes (Hazardous Waste). MDM will match and merge the record for Doe Run Herculaneum based on the business rules defined in the MDM engine by DNR. Once the master data record is matched and merged, MDM will insert a master data record for Doe Run Herculaneum that is being captured in MoEIS, MoCWIS, and Fees & Taxes. Any changes made to master data for Doe Run Herculaneum by the program applications will be made available by the MDM engine to other program applications. The master data record in MDM can also be used to generate reports or for sending to consuming applications, such as content management.



Figure ‑ MoGEM Future Data Flow

DNR’s planned approach includes using a master data management product, IBM Infosphere Master Data Management, to facilitate the master data management approach. Missouri’s vision includes architectural components and/or process steps to profile, consolidate, clean, and share master data.

DNR is progressing through a project schedule/roadmap that includes:

* Technical architecture determination including things such as tool selection and procurement, data model development, and data service development.
* Creation of data governance structure and processes.
* Business process improvement.
* Phased onboarding of DNR applications into MDM.
* Development of services including CROMERR compliant services and reporting.

### Business Processes

Discovery session Business Processes discussions covered a wide variety of topics (see Appendix D). Each discovery session section of this document will present the corresponding state’s responses in a table. Missouri’s responses are recorded in Table 3‑1. These notes are accurate as of the date of the Missouri discover session. Missouri is actively making progress on their MoGEM efforts and their processes are evolving as they move toward MoGEM implementation.

Table ‑ Missouri Discovery Session Business Process Information

|  |  |
| --- | --- |
| **Agenda Question** | **Notes** |
| Can you provide an overview of the facility data entity universe? | Data entity topics were discussed throughout discovery session meetings and will be addressed in the Data Model section (3.3.2.5) related to each discovery session. |
| Who are your main stakeholders/customers for master facility data? | DNR stakeholders include:   * Internal staff * External regulated entities * General public * Co-regulators (ex. EPA) |
| When do you perform data reconciliation and correction? | At the time of the discovery session, data reconciliation timing was not completely planned. The likely approach will be to reconcile at the end of the process, with a desire to get to the point where master data can be presented to users during processes. Initially, reconciliation will occur as data is flowing into the master data repository. |
| When, if ever, is master data presented to users in your systems? | Initially, reconciliation will occur as data is flowing into the master data repository. Presentation of master data to users in systems may be a future goal for inclusion in DNR applications. DNR noted that presenting data ‘up-front’ is a goal but requires application changes. |
| How do you handle Facility ID assignment (master and program-specific)? | DNR applications are responsible for their own ID assignment related to facility data. MDM will handle ID assignment to the master records. Source systems may require modification to closely tie source system IDs to MDM IDs. |
| Do you currently share facility data with EPA or other agencies? | DNR shares data with EPA where required by delegation authority. DNR is not flowing data to EPA FRS at this time. |
| What are your facility data management staffing approaches? Do you have any dedicated staff? | DNR does not have ‘dedicated staff’ but informally has staff fulfilling data stewardship roles. As the MoGEM project moves forward, DNR will continue to evaluate its approach to staffing and roles related to master data management and data stewardship. |
| What formal processes, operating procedures, and best practices do you utilize related to reconciliation and sharing/integration of facility data? | DNR does not currently have defined processes and SOPs related to reconciliation and sharing/integration of facility data. Historically, DNR systems have not done much sharing of data externally and internally. Where sharing is occurring, it is most likely in the form of reports. The ongoing MoGEM project and associated roadmap is beginning to address some of these areas.  MoGEM will include business rules that cover scenarios where, if certain rules are met between the data in program applications and MDM, automatic merging can occur. |
| What governance process do you use to determine how data is shared, integrated, and managed? | The Governance Board formed with the MoGEM effort is developing what these types of processes are going to be for DNR. Eventually the Governance Board will develop the processes and define standards for data integrity and quality. |
| Do you have a retention schedule? Is there ever a point at which you no longer ‘count’ these regulated entities and how are those data managed? | This topic was barely touched on and focused on formal retention schedules and policies. Most relevant discussion of ‘history’ occurred in the next discussion topic (associated specifically with facility closure and significant data modifications). |
| How do you manage a facility that closes or where key information, such as owner, changes? | The IBM Infosphere Master Data Management tool has the ability to log history. The tool refers to this is ‘survivorship.’ Facility closure denoting is also expected to be handled with out-of-the-box Infosphere capabilities. |
| How often do you attempt to compare state data with federal facility data? | DNR does not compare data with federal facility data regularly. DNR sends data through CDX/EN to programmatic systems at EPA and is aware that there are some significant data quality issues at a detailed level. |

### Data Model

The MoGEM *Technology Roadmap and Architecture Plan* and *Future Vision* documents note the importance of creating a well-planned and flexible data model to support the success of the DNR’s master data management efforts. MDM vendors evaluated by DNR did not include models for environmental programs. The logical data model created by DNR for MoGEM defines the following domains (definitions are from the *MoGEM Logical Data Model v1.1 FINAL* document):

* Organization Domain – The Organization Domain stores data for an organization. An ‘organization’ is a company, government entity, or non-profit responsible for activities at one or more associated Regulated Entities or related to one or more Persons.
* Person Domain – The Person Domain stores data for a person. A ‘person’ is an individual responsible for activities at one or more associated Regulated Entities or related to one or more Organizations.
* Regulated Entity Domain – The Regulated Entity domain stores data for a regulated entity. A ‘regulated entity’ is a facility, site, unit, or business that is regulated by DNR.
* Regulated Person Domain – The Regulated Person domain stores data for a regulated person. A ‘regulated person’ is person that is licensed or regulated by DNR.
* Program Interest Domain – The Program Interest domain stores values used for program interests associated to regulated entities and persons.

Figure 3‑2 depicts the MoGEM’s vision for the relationship between domains.

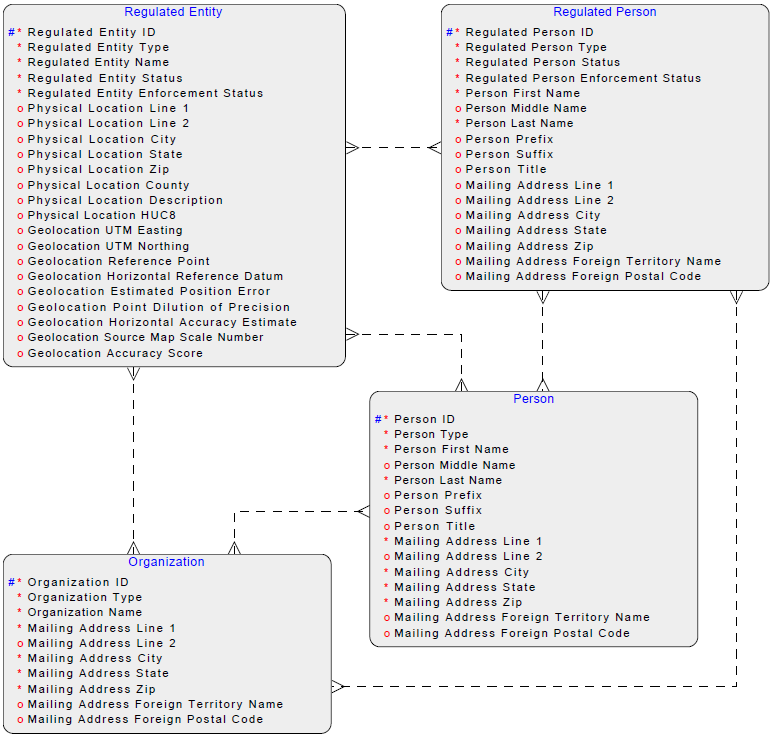


Figure ‑ MoGEM Domain Model

### Lessons Learned and Challenges

Given Missouri’s maturity status of being in the planning phases of facility master data management at the time of the discovery session, lessons learned or challenges discussion was a less significant part of the conversation. Some of the challenges that DNR faces as it moves forward with MoGEM were identified:

* MDM vendors that DNR evaluated do not have existing models for environmental data. Models to work from already exist in some domain areas. DNR had to define its own domains and data model for environmental facility MDM.
* The number of program interests being evaluated and put on the roadmap for integration into MDM is a significant challenge for DNR. In order to group integration into planned phases, DNR evaluated nearly 80 program interests for, among other things, technical readiness, business readiness, and business importance.
* DNR has identified program interests that are not currently in a ‘modernized platform’ for integration into MDM. These program interests are identified for later phases as they pose the challenge of requiring that they be modernized prior to integration with MDM.
* DNR faces challenges with staffing to schedule and perform the required work related to the phased integration of their program interests.

## Wyoming Discovery Session

The Wyoming discovery session was held by conference call on November 16, 2015. The following materials were provided by Wyoming either prior to or shortly after the discovery session.

| **Wyoming Discovery Session Materials Provided** |
| --- |
| Project Overview—WY DEQ Facilities Reconciliation-June2012  WY Site Profiler—ERD-CME  WY Site Profiler—ETL PROCESSES notes  WY Site Profiler—Design Specifications FINAL 9AUG13  WY Site Profiler—ERD-Golden Area  WY Site Profiler—nSITE ERDs  WY Site Profiler—Update to Director 8-26-13  WY Site Profiler—Technical Architecture Document—Final  Envl Interests SPCME |

### Participants

| **Wyoming Discovery Session Participants** | |
| --- | --- |
| Ben Way, WY DEQ  Kimberly Hoke, MO DNR  Lydia Lyshevski, CGI Federal  Ron Johnson, INDUS  Dave Smith, EPA OEI | Lee Kyle, EPA OEI  Lucas Gentry, CGI Federal  Jimmy Mahady, Ross Strategic  Matthew Kelly, EPA OEI |

### Level of Experience/History

Wyoming Department of Environmental Quality (DEQ) has had a facility master data management effort underway for approximately four years. DEQ contracted with Windsor Solutions to perform a facility data management assessment. Windsor’s Site Profiler application was implemented to pull data from ten source systems and perform data reconciliation at the back end of the programmatic business processes.

Past attempts to standardize facility and company information were unsuccessful due to cultural and technological challenges. An accumulation of issues pushed DEQ to pursue facility master data management:

* Increasing pressure to quickly provide senior management with cross-media information on all facilities within specific geographic areas of Wyoming. (In some instances, it would take DEQ weeks to provide the requested reports.)
* The need to provide comprehensive compliance history for a facility.
* The need to respond to citizen requests regarding facility information within a specific area.
* The need to support Site Remediation staff by providing facility information within the vicinity of a contaminated site.

### Technical Approaches

Wyoming’s primary technical approach is to employ the Windsor Solutions Site Profiler Resolution Application (known as SPRA or Site Profiler) to perform back-end reconciliation on data from ten programmatic databases. These ten databases contain the majority of facilities managed by Wyoming DEQ. DEQ’s approach is evolving with the Air Quality Division’s (AQD) new IMPACT data system which shifts the responsibility for facility data management in the hands of facility operators and other front-end users. With Site Profiler and IMPACT, DEQ is moving toward a hybrid approach to facility data management and reconciliation.

#### Site Profiler

The basic process related to Site Profiler for Wyoming DEQ is as follows:

1. The first ETL process executes nightly. The process pulls relative data fields from each of the ten source systems and moves the data into the Site Profiler staging environment.
2. A second ETL process executes and creates the presentation that will be shown to data stewards for review purposes.
3. These back-end processes perform basic reconciliation using algorithms to identify potential duplicate records. (The algorithm primarily involves facility name, location, and ownership data.) If found, the data steward/manager is notified that a potential set of duplicate facility sites exists and requires reconciliation.
4. The data steward/manager performs a manual reconciliation. The steward/manager can choose to merge the potential duplicates into a single ‘facility-site.’ The data steward/manager chooses the data (name, location, address, etc.) for the merged facility-site.
5. Data manipulation in the Site Profiler does not update information in the source systems. The data steward/manager must go back into the source system(s) to make corrections after the merge process. While correction of data in source systems is left to each source program, there is a push from upper management to continually increase data quality through practices like consistent standards and naming conventions.

#### IMPACT

IMPACT, the new AQD data system, is designed to perform front-end reconciliation and capture data entry and correction by the industry ‘data owners.’ Currently, IMPACT only handles AQD facility-site data and is not integrated as a source system to Site Profiler.

### Business Processes

Discovery session Business Processes discussions covered a wide variety of topics (see Appendix C). Each discovery session section of this document will present the corresponding state’s responses in a table. Wyoming’s responses are recorded in Table 3‑2.

Table ‑ Wyoming Discovery Session Business Process Information

| **Agenda Question** | **Notes** |
| --- | --- |
| Can you provide an overview of the facility data entity universe? | Data entity topics were discussed throughout discovery session meetings and will be addressed in the Data Model section (3.3.5) related to each discovery session. |
| Who are your main stakeholders/customers for master facility data? | DEQ stakeholders are primarily:   * Internal staff * External regulated entities (IMPACT) * General public |
| When do you perform data reconciliation and correction? | The Site Profiler Application pulls data from ten source systems and performs data reconciliation at the back-end (see Section 3.3.3.3). The ETL processes execute nightly, so data corrections entered in source systems are frequently updated in the Site Profiler.  AQD’s IMPACT application is designed for front-end reconciliation. Facility data is entered and corrected at the front-end by industry ‘data owners.’ IMPACT only handles AQD facility-site data and is not yet connected as a source system to the Site Profiler. The approach for IMPACT integration with Site Profiler is still under investigation at DEQ. |
| When, if ever, is master data presented to users in your systems? | The ten source systems that feed Site Profiler do not present master data to system users.  IMPACT presents master facility data to the front-end users for verification and correction. The responsibility of data entry and correction is shifted to industry but some controls will be maintained by DEQ. |
| How do you handle Facility ID assignment (master and program-specific)? | Facility-site IDs are assigned at an abstract facility level (ex. an oil refinery level). The IDs are assigned by each data management system. As part of the data integration process, the Site Profiler application assigns its own Site Profiler ID to the reconciled facility-site record. Site Profiler maintains and publishes all system IDs associated with a facility record.  Unit-level IDs may be assigned depending on the program data system. Unit-level IDs are not included in the Site Profiler. |
| Do you currently share facility data with EPA or other agencies? | DEQ shares data with EPA through CDX/EN data flows. |
| What are your facility data management staffing approaches? Do you have any dedicated staff? | The programs that own the ten source Site Profiler data systems provide data stewards/managers who are responsible for reconciling potential duplicates identified by the Site Profiler ETL processes. |
| What formal processes, operating procedures, and best practices do you utilize related to reconciliation and sharing/integration of facility data? | Formal processes and operating procedures were not discussed. If relevant, the IPT may schedule a follow-up during a subsequent phase to obtain this information. |
| What governance process do you use to determine how data is shared, integrated, and managed? | Formal processes and operating procedures were not discussed. If relevant, the IPT may schedule a follow-up during a subsequent phase to obtain this information. |
| Do you have a retention schedule? Is there ever a point at which you no longer ‘count’ these regulated entities and how are those data managed? | This topic was barely touched on and focused on formal retention schedules and policies. Most relevant discussion of ‘history’ occurred in the next discussion topic (associated specifically with facility closure and significant data modifications). |
| How do you manage a facility that closes or where key information, such as owner, changes? | IMPACT has a sophisticated versioning system. When an emission inventory is submitted, the current facility information is ‘locked’ to the inventory that was submitted. DEQ hopes that fewer errors will be encountered when sending XML data to EPA. |
| How often do you attempt to compare state data with federal facility data? | DEQ did not indicate that regular comparison between state and federal data is performed. |

### Data Model

Wyoming requested that Site Profiler data models are not shared with the Facility IPT materials. Direct specific questions related to the data model to Wyoming DEQ IPT members.

In general, the Site Profiler logical data model contains the following primary entities:

* Facility
  + Related to 0 or more Environmental Interests.
* Environmental Interest
  + Related to 1 and only one Facility.
  + Related to 0 or more Entity records through an affiliation table.
* Entity
  + Related to 0 or more Environmental Interests through an affiliation table.
* Additional entities related specifically to Compliance, Monitoring, and Enforcement activities.
* Additional entities related to ETL, security, and administration functions.

DEQ indicated that complex corporate structures may not be fully evident in the Site Profiler data through the Facility to Environmental Interest to Entity relationships.

The Site Profiler data stores approximately 56,000 abstract facility-site records and some top-level ‘scrape’ of compliance, monitoring, and enforcement data.

### Lessons Learned and Challenges

Lessons learned and challenges specific to Wyoming DEQ experiences were identified during the discovery session:

* Field offices and field office databases operate autonomously, largely because of the communications infrastructure making it difficult to work over the Internet.
* It is challenging to determine how to integrate IMPACT with Site Profiler given that IMPACT performs front-end data management and reconciliation. (For example, when integrated with Site Profiler, DEQ needs to determine whether reconciliation notices/requests will be routed to DEQ data stewards/managers or back to the regulated users).
* Complex corporate data structures are difficult to report on from Site Profiler.
* Some facility-site data is sensitive. There is concern over publishing abandoned mine land information. In these cases, the coordinate information for abandoned mine facility-sites are truncated to avoid disclosing specific locations.

## New Jersey Discovery Session

The New Jersey discovery session was held by conference call on November 13, 2015. The following materials were provided by New Jersey either prior to or shortly after the discovery session.

| **New Jersey Discovery Session Materials Provided** |
| --- |
| Site Masterfile SOP & Standards for Data Entry |

### Participants

| **New Jersey Discovery Session Participants** | |
| --- | --- |
| Jim Bridgewater, NJ DEP  Kimberly Hoke, MO DNR  Ron Johnson, INDUS | Theresa Pagodin, NJ DEP  Lee Kyle, EPA OEI  Lucas Gentry, CGI Federal |

### Level of Experience/History

New Jersey is a mature state in terms of environmental facility master data management. New Jersey Department of Environmental Protection’s (DEP) facility master data management is a component of a broader business process reengineering and systems development effort that was initiated in 1994 with the goal of better serving the regulated community and the public. One of the results of that effort was a recommendation to consolidate the many application and database silos into an enterprise-wide environmental management system. This system, eventually called DEP’s New Jersey Environmental Management System (NJEMS), was conceived to include a master data management component to define unique regulated ‘sites’ and the regulatees at those sites. This structure was intended to provide DEP the ability to evaluate and analyze all environmental factors related to a specific site. This master data management component, the Site Masterfile, went into production at DEP in 1997. DEP actively continues to migrate existing and new regulatory programs and business processes into the Site Masterfile and NJEMS, in part to enforce master data management practices across the agency.

The common challenges faced by organizations with multiple data silos were the driving forces behind the vision and implementation of DEP’s Site Masterfile and NJEMS systems. Facility master data management capabilities of Site Masterfile and NJEMS mitigate those challenges. With these tools and continuous improvement, DEP is able to provide a clearer and more holistic view of a location or ‘site.’

### Technical Approach

At the core of DEP’s environmental management approach is the DEP– and CGI-developed NJEMS/TEMPO, Site Masterfile, and Regulatory Systems Portal (RSP) solution. Facility management is performed in the Site Masterfile functionality of NJEMS. As shown in Figure 3‑3, regulated entities and DEP staff interact with DEP through the RSP. DEP staff can perform back-end capabilities, including facility management, directly within NJEMS.

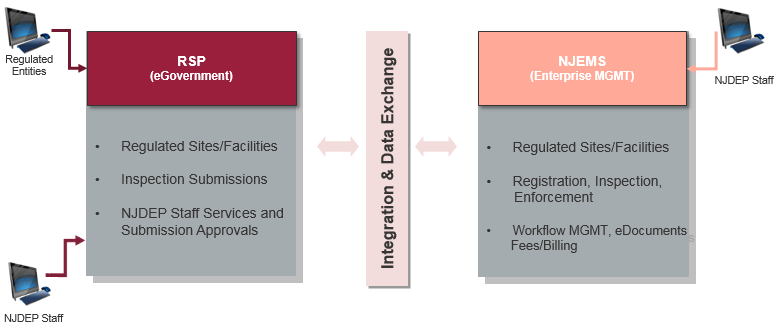


Figure ‑ NJDEP Architecture

Facility data management within DEP is a shared, or core, capability within NJEMS. NJEMS is built around a ‘core’ and ‘plug-in’ model, as shown in Figure 3‑4. Programmatic capabilities are plug-ins and take advantage of core capabilities. Core and plug-in software and data are tightly integrated.

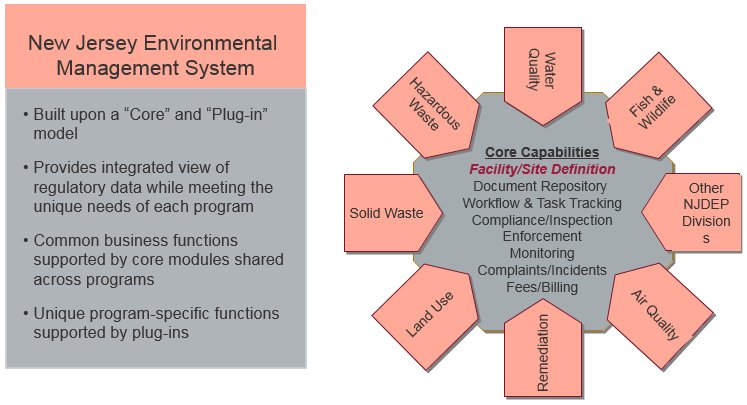


Figure ‑ NJEMS Architecture

In addition to these core architectural components, DEP uses the following tools:

* HighView (CACI) – Imaging and content management
* E2 (EnfoTech) – Ambient monitoring data
* NJGeoweb (ESRI) – Geospatial
* SDWIS (SAIC/USEPA) – Drinking Water
* COMPASS (CGI) – Sampling/analytical laboratory data
* DataMiner (DEP) – Reporting application built on a Business Objects framework that makes much of DEP’s data publically accessible

While DEP does not publish master data via web services, it regularly refreshes a Site/Program Interest GIS layer that is used by both internal and external NJGeoweb application users.

### Business Processes

Discovery session Business Processes discussions covered a wide variety of topics (see Appendix C). Each discovery session section of this document will present the corresponding state’s responses in a table. Wyoming’s responses are recorded in Table 3‑3.

Table ‑ New Jersey Discovery Session Business Process Information

| **Agenda Question** | **Notes** |
| --- | --- |
| Can you provide an overview of the facility data entity universe? | Data entity topics were discussed throughout discovery session meetings and will be addressed in the Data Model section (3.3.4.5) related to each discovery session. |
| Who are your main stakeholders/customers for master facility data? | DNR stakeholders include:   * Internal staff * External regulated entities * General public * Co-regulators (ex. EPA) |
| When do you perform data reconciliation and correction? | Generally, data reconciliation and correction is performed at the end of the process. The systems include some built-in procedures and controls (such as forcing three searches for a master entity prior to allowing the creation of a new entity) aimed at maintaining data quality. DEP has dedicated staff that reviews master data changes and performs quality assurance. |
| When, if ever, is master data presented to users in your systems? | Internal users have the ability to see all data. Once a Site record has gone through the quality assurance processes, only data administrators are able to make changes. Internal users can request changes that are routed through the quality assurance processes. Program Interest data is managed by the program area and can be changed by program users with a specific security role. Master data is presented throughout the system.  External users, through the RSP, are presented with Site, Program Interest, and Organization data as it appears in the master data component (the Site Masterfile). Some online services allow direct modification of certain data elements (for example, mailing address and contact information), but other data changes made by external users are routed through quality assurance processes before being accepted as master data.  Some online services make use of embedded maps that allow a location can be selected and coordinates to be entered into the system. Additionally, standard lists are used for some fields such as County and Municipality to normalize data and control quality. |
| How do you handle Facility ID assignment (master and program-specific)? | Unique IDs are assigned to all master entities (ex. Sites, Program Interests, Organizations, Subject Items, and Regulated People). Site IDs are system-generated. Program Interest IDs can be system-generated or entered by the program. Alternative IDs can be assigned to Program Interests. |
| Do you currently share facility data with EPA or other agencies? | DEP sends all Site and Program interests ID information to FRS. |
| What are your facility data management staffing approaches? Do you have any dedicated staff? | DEP has two people/full time equivalents (FTEs) dedicated to quality assurance and quality control (QA/QC) of master data. |
| What formal processes, operating procedures, and best practices do you utilize related to reconciliation and sharing/integration of facility data? | DEP provides training, has system-enforced rules, and uses regular procedure reminders to enforce data quality and data entry standards. Duplication of master data, particularly Sites, is still a common issue that requires manual QA/QC and correction. |
| What governance process do you use to determine how data is shared, integrated, and managed? | The primary governance mechanism is through review of master data and master data modifications by dedicated QA/QC staff. |
| Do you have a retention schedule? Is there ever a point at which you no longer ‘count’ these regulated entities and how are those data managed? | This topic was barely touched on and focused on formal retention schedules and policies. Most relevant discussion of ‘history’ occurred in the next discussion topic (associated specifically with facility closure and significant data modifications). |
| How do you manage a facility that closes or where key information, such as owner, changes? | DEP can exclude master data based on status flags/indicators or through other filtering mechanisms. NJEMS maintains a significant history.   * At the Site level, a history of Site Name and Property Owner can be maintained with start and end dates. * At the Program Interest level, there is an ‘Active’ flag, start and end dates, and the responsible entities for the Program Interest. * A Program Interest can have a start and end date history.   In addition, some program functionality maintains snapshots of master data associated with specific ongoing or historic environmental actions taking place at a Site or Program Interest. |
| How often do you attempt to compare state data with federal facility data? | DEP sends data to FRS but does not undergo efforts to compare state data to federal facility data. |

### Data Model

The NJEMS logical data model consists of the following primary entities:

* Site – A geographic entity of interest in the system.
* Program Interest (Facility) – An entity (located at a Site) which is of interest in the system, typically because the entity is subject to regulatory oversight.
* Organization – An entity (that is not a person) that is in relationship with other entities in the system. An organization can be a company, utility, association, federal/state/local agency, or any other non-person that is part of a relationship.
* Subject Item – A portion, unit, operation, or other aspect of a Program Interest that is of interest in the system, often because it is or has been subject to one or more enforceable requirements.

The following are not ‘master’ entities, but their relations are important in the NJEMS logical data model, as shown in Figure 3‑5.

* Portal User – The identity that the RSP users to login and interact with the portal.
* Agency Contact – A staff member contact at an entity (ex. Organization or Program Interest) where agency and non-agency contacts are treated separately.
* Non-Agency Contact – A contact that is not a staff member.



Figure ‑ NJEMS Logical Data Model

The NJEMS logical data model starts at the Site level. Site has a geographic or locational connotation; it is defined by block and lot, if available, as well as physical address and coordinate data, if available. Sites have a name and an ability to track name history. A history of property owners can also be preserved with the Site. Program Interests are linked to Sites. Subject Items are linked to Program Interests. Sites, Program Interests, and Subject Items are all assigned unique identifiers.

The relationships between Sites and Program Interests can be used to model complex operations. Sites can be cross-referenced to create an even larger representation such as a remediation activity.

The terminologies used are shared consistently within DEP but not across other State agencies. An online glossary is available to users accessing DEP data through DataMiner.

### Lessons Learned and Challenges

Lessons learned specific to New Jersey DEP experiences were identified during the discovery session:

* Business process reengineering should be the first step in moving toward a master data management approach. It is inefficient to modernize bad processes.
* Communication between programs is essential. There needs to be one authority that can resolve conflicts. Data quality is integral to DEP’s functions and must be a top priority.
* Management (particularly mid-level and supervisor management) has to be onboard and must accept the level of effort that is required to maintain master data. Management must recognize that efficiencies will be realized if the up-front effort is consistently applied.
* Consistent management and enforcement of the policies and procedures related to accurate and timely input and maintenance of master data is essential to success. Some DEP programs include metrics related to proper and correct use of master data into employee performance assessments.

The following challenges specific to New Jersey DEP experiences were identified during the discovery session:

* Sometimes, even with mature guidance and procedures, conflict arises within programs. Reference data and configuration procedures can often resolve conflicts but there are occasions where programs will attempt to intentionally create a duplicate Site.
* As programs migrated their data into DEP’s master data, one of the most difficult features to establish was the Program Interest numbering scheme. Accommodations were made to continue the use of some existing numbering conventions, but it was difficult for programs to accept that new identifier assignments would be made by the system automatically.
* An ongoing challenge for DEP is that People do not fit well into the NJEMS logical data model and system. DEP continues to explore options to better accommodate the tracking of people as master data.

## EPA/FRS Current State and Future Plans

EPA recognizes that the concepts around ‘facility’ are foundational to E-Enterprise. Continuous improvement of all partner tools associated with the capture and maintenance of facility data will allow E-Enterprise to provide a consolidated view of what is occurring at a facility. Moreover, these improvements will provide consistent facility information to the public, regulated entities, and co-regulators interacting at different touch points across E-Enterprise.

### Known FRS Issues

The FRS team has identified the following issues and challenges to be addressed to meet current needs and to align with E-Enterprise:

* FRS was built to integrate data at the ‘end of the pipe’ rather than at the source. This design has proven to be complicated, expensive, and to result in changing of IDs that makes reporting and establishing the history of an entity very complex.
* FRS has been focused at the facility level and needs to be more granular. FRS needs to accommodate large ‘facilities’ that are made up of multiple regulated components, each with a separate and persistent ID.
* ‘Facility’ is defined differently across different regulations. A new lexicon that is independent of the regulations is needed.
* FRS does not have adequate capabilities to maintain facility and component history.

### FRS Issue Resolution Approach

Collaboration with partners across E-Enterprise is one of the pillars of EPA’s approach to facility and E-Enterprise. This is accomplished through efforts such as the Facility IPT. The following are the primary components to EPA’s approach:

1. Data Stewardship – As a component to the success of facility management within E-Enterprise, EPA envisions a network of data stewards provided with data quality tools and fully understood business requirements. The FRS data stewardship tool, the Facility Linkage Application (FLA), is in use by more than 140 data stewards at EPA, states, and municipalities. As facility data integration matures, data authority may move to SLTs. The exact nature of data stewardship, use of the FLA or other data stewardship tool, and the supporting resources that EPA will need to provide to SLTs will be explored in Phase II. EPA is currently developing a major release of the FLA to provide enhanced capabilities and to work within the new FRS data model.
2. High-Quality Data Entered at the Source – EPA views collaboration with regulated entities in order to capture high-quality data from the source as essential to the future success of facility management within E-Enterprise. The Facility IPT explores concepts related to capturing data at the source throughout the enterprise. EPA continues to work within the constructs of CDX and FRS to provide services to capture facility data up-front from direct reporters to EPA. See Appendix E for an analysis on the significant burden reduction benefits gained for TRI direct reporters.
3. Harmonize Core Facility Data Elements – In order to achieve the sharing of core facility information across E-Enterprise, partners must achieve a reasonable level of data element harmonization. EPA continues to modernize the FRS into a flexible model that will enable synchronization with partners.
4. Develop and Implement Widely-Utilized Facility Identifiers – To support the sharing of facility ‘profile’ information for E-Enterprise every facility or regulated component will be required to have persistent identifiers.

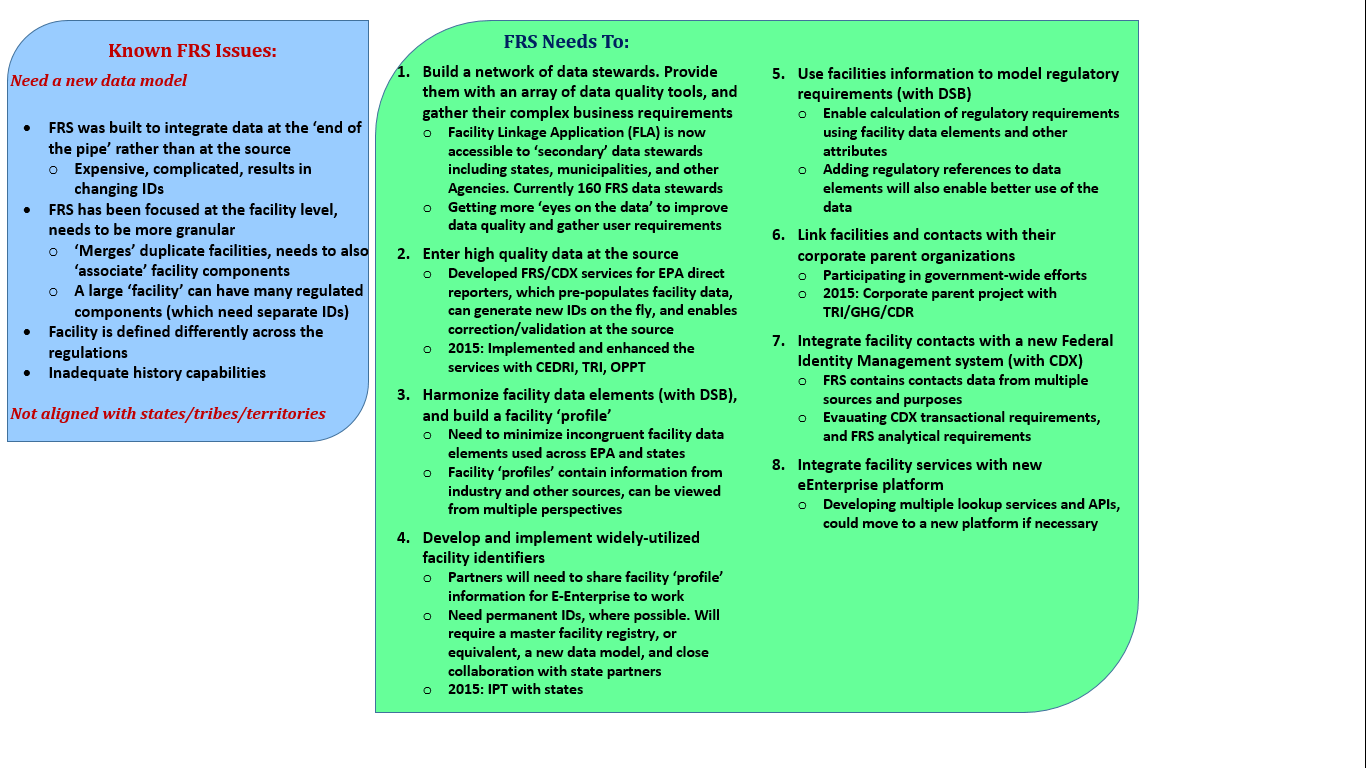
In the FRS data model being replaced, facility IDs can change. This causes issues for reporters and regulatory agencies. Capturing high-quality data at the source by connecting users to master records corrects the problem.

1. Use Facilities Information to Model Regulatory Requirements – Enabling the systems to calculate regulatory requirements based on facility data elements will help E-Enterprise partners to provide better service to the regulated community.
2. Corporate Parent Relationships – In order to achieve the goals related to E-Enterprise, EPA views the ability to establish relationships between facility data and corporate ownership as an essential component. EPA’s ongoing modernization of the FRS data model will support the linking of contacts (people or organizations) with their facilities and corporate parent organizations.
3. Integrate Facility Contacts with a New Federal Identity Management system – In order to better manage identities and their interaction with facility data, EPA and FRS will need to manage contacts as master data.
4. Sharing Core Facility Information – EPA views collaboration and sharing of a core facility ‘profile’ with partners/co-regulators as a critical component to facility management within E-Enterprise. Among other capabilities, this will allow E-Enterprise to provide complete, current, clear, and consolidated facility information to regulated entities (i.e. a reporter at a facility). Ongoing collaboration with state partners through the Facility IPT is an essential component to achieving these goals.

The IPT envisions a common mechanism for sharing facility data that may initially leverage current FRS investments and seeks to utilize data from SLTs that has already been integrated at the SLT level. Collaboration also informs the ongoing modernization of the FRS data model. To position EPA to share data and services with partners, EPA must maintain a flexible data model.

Note: FRS is EPA’s tool that aims to integrate facility information at EPA. FRS is not being suggested as the master place to store all integrated facility data across the enterprise. In this instance, FRS is treated as a single state MDM repository.

Known FRS issues and proposed approaches are summarized in Figure 3‑6.

Figure ‑ Known FRS Issues and Proposed Approaches

### Data Model

The FRS data model organization creates a flexible architecture that will enable EPA to contribute to a shared facility ‘profile’ across E-Enterprise. The EPA approach can be described as a ‘thing’ model, as shown in Figure 3‑7. ‘Things’ are objects of interest/facilities; they can include regulated and non-regulated business processes and interests such as emission points, boilers, storage facilities, mobile sources, sampling points, and so on. All ‘things’ are assigned permanent IDs.

Through point locations and polygons, EPA will have the capability to capture where ‘things’ are at a detailed level. The model will be able to support interest relationships at complex environmental sites such as military bases, airports, and industrial campuses.

The model will enable the capture of detailed attributes about ‘things’ including equipment/process details, industry codes, contact IDs, parent company IDs, and additional attributes.

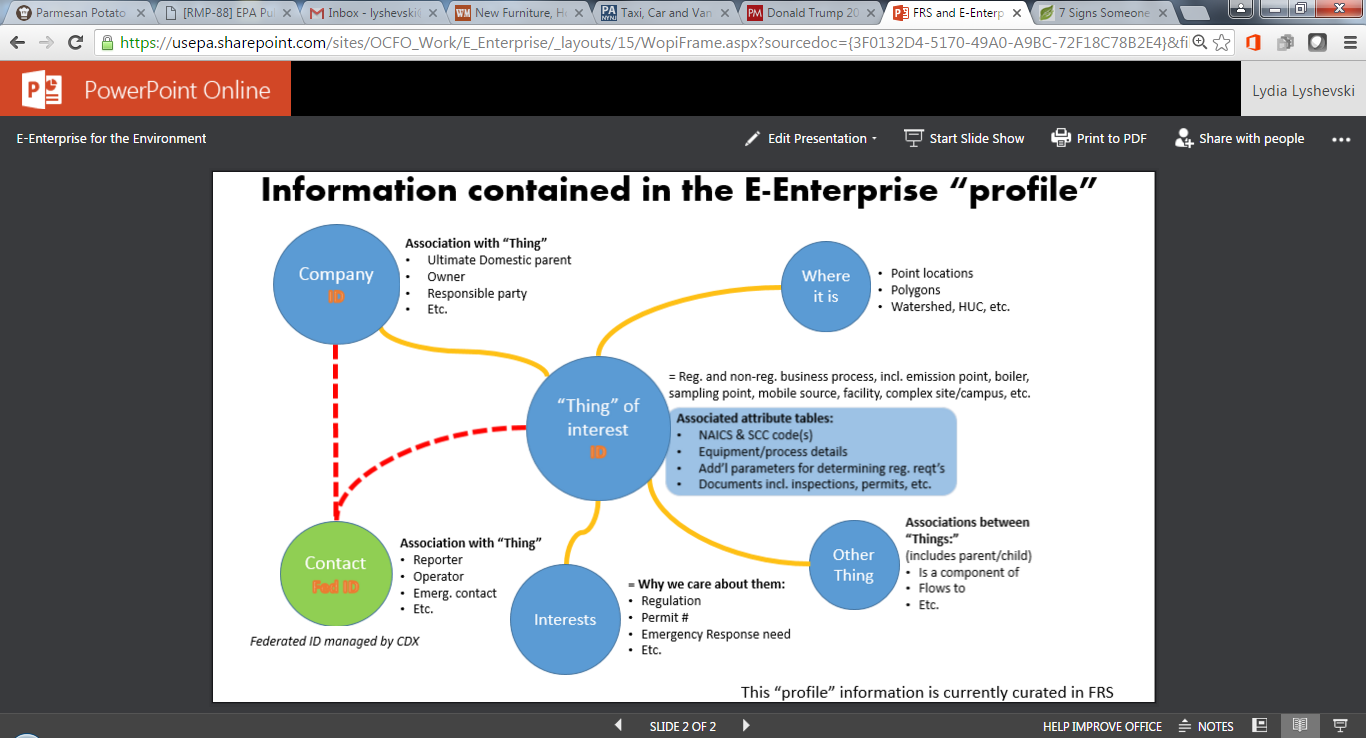


Figure ‑ FRS ‘Thing of Interest’ Model

Essential to modelling complex interests, the EPA model will also support flexible associations between ‘things.’ This will enable the capture of relationships such as:

* Thing 2 (ex. boiler) is a component of thing 1 (ex. plant)
* Thing 3 (ex. waste stream) flows to thing 4 (ex. outfall)
* Company X owns thing 1
* Contact A is from Company X

## Summary

The discovery sessions provided an opportunity to share information between the parties and to explore some aspects of facility management at a detailed level. The sessions highlighted emerging themes from the IPT as a whole. These sessions reinforced that ‘facility’ is important to individual partners and E-Enterprise as a whole, that facility management is challenging, and that efforts to come together around facility in E-Enterprise must focus on areas where partners have the most in common.

The following sections identify the key summary-level takeaways from the discovery sessions.

### Maturity Level

The three discovery sessions intentionally included participants at different facility master data management maturity levels. Missouri is moving forward quickly but is still in the ‘Getting Started’ phase. Wyoming has been involved in facility master data management for approximately four years and is embarking on new capabilities through their AQD’s IMPACT system. New Jersey has made facility master data management a cornerstone of their integrated systems for approximately twenty years.

Section 3.2 presents information gathered during an IPT member survey. The results show that the IPT members are spread out across the spectrum of experience related to facility master data management.

***As E-Enterprise and the work of this Facility IPT move forward, the vision and next steps identified must continue to take into account the fact that the solution(s) for sharing facility within E-Enterprise must provide utility to, and be interoperable with, partners who are taking a range of integration approaches and who are also at various levels of facility master data management maturity across programs.***

### Technical Approach

Technical approaches to facility master data management vary. Not only do the approaches and tools differ between partners, but approaches can even vary within an agency. Broadly speaking, the IPT has discussed the general approach to master data management in terms of:

* Back-end or ‘end of process’ reconciliation
* Up-front reconciliation
* Hybrid reconciliation approach

Missouri reported that its current plans are to initially perform reconciliation at the end of the process, with a desire to get to a point where master data can be presented to users during processes.

Wyoming reported a back-end approach for ten data systems flowing to their Site Profiler; at the same time, through their new IMPACT system, Wyoming also has an approach that puts the burden of data entry and correction on the regulated entity.

New Jersey takes a unique hybrid approach where programmatic systems are directly tied to, and share, the facility master data. In some senses, New Jersey is performing up-front reconciliation in that a new entity must be created in the master data repository prior to a program taking action with the entity. In some processes New Jersey is clearly performing back-end reconciliation where changes to data proposed by a facility user through the RSP interface or by staff members through master data change requests, are reviewed and approved by dedicated quality assurance staff members before they are entered into/updated in the master data repository.

Through modernization of FRS, EPA takes a hybrid approach as well. EPA seeks to shift the responsibility of data entry and correction up-front in the process on the regulated entity. Meantime, EPA also relies on a network of data stewards to ensure data quality and integrity.

It is important to note that there will be a variety of tools and utilities to perform facility master data management even if partners follow the same or similar approaches to those captured in this document. The 3 discovery session participants use 3 different tools as their primary mechanism for performing facility master data management:

* Missouri – IBM Infosphere Master Data Management
* Wyoming – Windsor Site Profiler (now known as nSite)
* New Jersey – CGI NJEMS/TEMPO Site Masterfile

While there is some overlap of tools being used within multiple states, this is by no means an exhaustive or representative list of the tools in use across potential E-Enterprise partners.

***As E-Enterprise and this Facility IPT move forward, the vision and next steps identified must account for the fact that partners will be performing their facility master data management processes at various points in the regulatory lifecycle. Additionally, each discovery session participant identified a requirement to perform back-end data stewardship. This commonality may indicate a continued need for data stewardship at the enterprise level within the future of E-Enterprise.***

### Data Model

Partner data models and approaches to the logical organization of data differ. The most common concept between models is some form of top-level, primarily geographically-based entity, at which a regulated activity or process is occurring. While there are differences in environmental regulation and approaches among partners, all of the Facility IPT partners are in fundamentally the same business and deal with the need to store data about similar real-world entities. This creates a situation where partners are tending to store similar information about things like boilers, outfalls, hazardous waste generators, etc. Approaches to how those individual units or interests are stored and whether or not they are stored as part of master data with unique identifiers vary from partner to partner.

***Discussion on the sharing of facility profile information within the Facility IPT has focused on concepts such as finding a ‘common enough’ data standard to allow participants to share facility profile information. IPT meeting discussion and discovery sessions have proven that there exist enough differences between partners that a ‘common enough’ standard will be necessary and should be defined as part of the next steps of this collaborative facility data management effort.***

### Commonalities

While level of maturity, technical approaches, and data models can vary significantly from partner to partner, the initial pain points that led Facility IPT discovery session participants to embark on facility master data management strategies were very similar.

* Typically, silo data systems posed significant challenges to an organization’s ability to perform holistic reporting required by the state government and the general public.
* All discovery session participants identified facility data as the most important class of data that enables their organization to provide holistic and accurate data outputs.
* Discovery session participants observed trends in government to provide better services to regulated entities. Facility master data management was identified as foundational to the ability to improve service.

Discovery session participants also noted several common challenges faced while implementing or maintaining facility master data management approaches. These common challenges ranged from technical challenges (such as having to deal with the proliferation of source systems based in various technologies) to ongoing data stewardship challenges (related to the need to reduce duplicate facility information) to staff resource challenges (partners must dedicate staff resources to enforcing data standards and ensuring good data stewardship).

***When working with their respective government oversight, industry, and programs to implement and maintain quality facility master data, our partners are often faced with similar pain points and challenges. Given that our partners’ various efforts are rooted in similar business processes and challenges, E-Enterprise should seek to establish enough common ground to put forward a facility data sharing solution that can be beneficial to members throughout the enterprise.***

### Considerations for Next Steps

The following topics, identified through discovery session analysis, should be considered when planning for subsequent phases related to E-Enterprise facility data management:

* It is possible that more mature partners are more suitable candidates for efforts such as pilots and prototypes of facility management services. While it is reasonable that encouraging less mature partners to get involved early might be a good approach, partners who are just getting started may be under significant project deadlines and funding constraints (making it particularly difficult for less mature partners to participate at a deep level). All willing participants will be considered, but the collective group may find that more mature members are in a better position to actively participate.
* Facility IPT discussions have been collaborative and focused on finding common ground (be it in technical approaches or data standards) in order to move forward with service-based facility management for E-Enterprise. While finding common ground will continue to be a vital aspect of these efforts, discovery sessions have made it clear that some significant differences will exist among partners and that explored approaches must be *flexible* in order to account for those differences.
* Despite noted differences in technical approaches and data models, it should be possible to develop a framework and a strategy to perform service-based facility data sharing through E-Enterprise. Because of these differences, establishing governance and trust between partners (see Section 4.3) will be especially important to the success of these efforts.

# Facility Data Management Partnership Vision

During the course of Phase I of this Facility IPT effort, a vision for E-Enterprise facility data management began to emerge. Facilitated IPT discussions related to ‘Problem, Opportunity, and Vision,’ ‘Key Hypothesis,’ and ‘Principles’ combined with planning and expertise brought by IPT partners form the basis of this endeavor. This vision will be more fully explored in subsequent phases of this effort.

## Users

The vision for E-Enterprise facility data management must address facility data needs for the following E-Enterprise user types:

* Co-Regulators – Users who are also members or staff of partner environmental regulators. Co-regulators may be state, tribe, federal or local government environmental regulators (ex. agency staff, both program and dedicated facility staff.)
* Regulated Users – A user who is acting on behalf of themselves, a company, or a facility that is subject to federal, state, tribal, or local environmental regulation.
* Public Users – A user who is a member of the public and is not acting on behalf of a regulatory agency or a regulated entity.

## Principles

Guiding principles were developed to highlight key ideas from IPT analysis, deep dives and group discussions. Facility Data Sharing Principles cover the essential characteristics of a system that supports facility data integration. These standards drive effective use of facility data in a manner that supports E-Enterprise objectives.

Thirteen principles, summarized in Figure 4‑1, form the foundation of effective facility data integration. Each item is described in greater detail below. Collectively, the principles will focus future conversations and guide subsequent phases of the Facility Data Integration IPT.

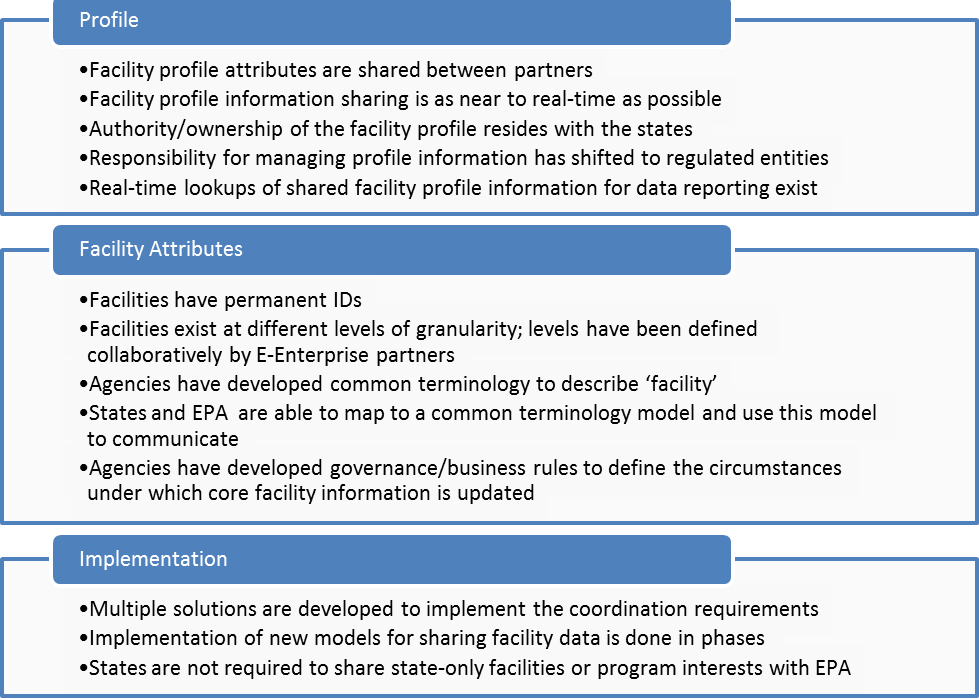


Figure ‑ Facility Data Sharing Principles Outline

The IPT had an overwhelmingly positive response to the first draft of the Facility Data Sharing Principles, as shown in Figure 4‑2. The outcome validated that the IPT has taken positive steps toward developing a common understanding of the vision and next steps required to ensure the success of Facility Data Integration nation-wide. The poll consisted of 27 participants, including meeting organizers who did not vote. Results were used to gauge the level of understanding, measure consensus, and identify gaps.

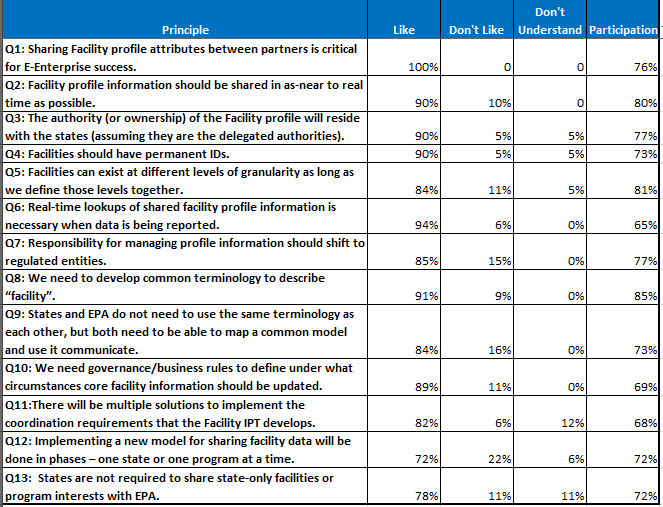


Figure ‑ IPT Response to Straw Facility Data Sharing Principles

### Facility Data Sharing Principles:

1. Sharing Facility profile attributes between partners is critical for E-Enterprise success.

All E-Enterprise partners (SLTs and EPA) need to know similar things about an entity. That can only work if agencies share core facility profile information. This principle acknowledges that E-Enterprise cannot be a success until partners determine how to share facility information. Sharing facility profile information could allow for the regulated community to report and access the same information from various E-Enterprise partner portals and applications.

Note: The IPT still needs to define what ‘core profile’ means. Varying levels of thought and effort have been performed by individual IPT members. Phase II will explore more fully what this means for E-Enterprise.

1. Facility profile information must be shared in as near to real-time as possible.

In the expanding and modern business of environmental stewardship, reliance on weeks– or months-old facility information will no longer be acceptable. If a partner (such as a state) updates a facility record or generates a new facility, the update needs to be communicated in as close to real-time as possible to the enterprise facility registry. If not, agencies will always be reconciling incorrect or outdated information and straining to stay abreast of the latest data.

1. The authority (or ownership) of the Facility profile will reside with the states (assuming that the states are the delegated authorities).

States with delegated programs manage more facility information than EPA does in EPA’s direct reporting systems. While this data resides with the states, states acknowledge that some of their facilities also report directly to EPA. This IPT has begun to explore some of the business rules and governance issues around this principle. Subsequent phases will explore more fully how E-Enterprise partners manage their data.

1. Facilities must have permanent IDs.

The sharing model falls apart if IDs are in flux. If a state ‘owns’ the data, then the state needs to determine how to create and communicate a master ID to EPA. Further, the state must retain the master ID for the corresponding facility. Administrative tools or utilities for maintaining IDs may be required.

1. Facilities can exist at different levels of granularity as long as agencies define those levels collaboratively.

Some programs define ‘facility’ at a lower level than other programs. (For example, 1 ‘facility’ for a water permit might represent 4 ‘facilities’ for the air program.) As long as the data model and data sharing procedures/services are developed with this concept in mind, implementing shared facility data can still work. This is a complex principle that requires further refinement in Phase II.

1. Real-time lookups of shared facility profile information are necessary when data is being reported.

When reporting is done, real-time lookups of the master facility record need to be available so that a user can easily identify their facility and continue with reporting. The goal is to avoid adding ‘new’ facility records for existing facilities. These real-time lookups will enable reporters to search for their facility, pre-populate reporting forms, and correct/validate data.

1. Responsibility for managing profile information must shift to regulated entities.

If a regulated entity requests a change to their profile information, information will not be changed automatically. States need to have the ability to vet the change. Moreover, regulated entities must have the ability to make an update or note the need for an update in their submission. Depending on the outcome of the state’s assessment, the update may not be applied to the master facility profile.

1. Agencies must develop common terminology to describe ‘facility.’

The IPT recognizes that the lexicon differs across locales and programs, in part due to definitions in state and federal laws and regulations. For this initiative to succeed, it is crucial to create a common, nation-wide understanding of ‘facility’ from a data integration standpoint. At minimum, data elements at the local level should map across E-Enterprise partners.

1. E-Enterprise Partners do not need to use identical terminology within their agencies, but all partners must understand the shared definitions and must be able to map to a common model and use this model to communicate.
2. Agencies need to develop governance/business rules to define the circumstances under which core facility information should be updated.

Updates do not apply to the master ID, which needs to remain permanent. This concept includes rules for scenarios that result in updates to the master profile record (for example, the procedure for updating the facility name, initiated either by the facility or the state).

Governance is addressed at a high level in Section 4.4 and will be explored in greater detail in subsequent phases of the IPT. These principles are expected to be refined and enhanced in future phases as the facility data integration initiative matures.

1. There will be multiple solutions to implement the coordination requirements that the Facility IPT develops.

Distinctive needs of each organization need to be recognized. While some agencies may achieve alignment via small tweaks, high levels of effort may be required to achieve alignment at other organizations. Solutions can emerge from multiple sources – a Commercial off-the-shelf (COTS) vendor, state-developed MDM, FRS APIs (application program interfaces) for facility profile data management, and so on.

An API is a set of data structures, protocols, routines and tools for accessing a web-based software application. They can be used to facilitate real-time data sharing between state and EPA systems.

1. Implementing a new model for sharing facility data will be done in phases – potentially one state or one program at a time, or several parallel prototypes at a time.

Given each agency’s unique needs, solutions must be flexible enough to accommodate a wide spectrum of organizations. The IPT aims to learn from each implementation and apply these insights to future implementations. Parallel prototyping with several partners may be viable via an Agile approach. Prototypes will target a variety of scenarios (ex. different states, programs, durations, and system maturity levels) to help prevent oversights.

1. States will not be required to share state-only facilities or program interests with EPA.

States and regions will continue to have the option to exchange data via the CDX network. Facility data relevant for E-Enterprise reporting is expected to be merely a subset of the full scope of facility data relevant at SLT levels.

## Components

The following sections introduce a draft set of components that will help form a vision that will be explored further in subsequent phases of this effort. These components were introduced during the 2015 Exchange Network National Meeting (EN2015) and were further discussed in subsequent Facility IPT meetings. The components will help form a starting point for this effort’s next phases.

### API First

Quality APIs that are flexible enough to work with multiple partner systems are expected to be essential to the success of sharing facility information within E-Enterprise. API-First approaches are becoming more commonly used in software and service development. The approach may be especially useful in the context of this E-Enterprise effort given that the nature of API-First benefits align with aspects of this effort.

Some API-First approach benefits that are applicable to this effort include:

* Separation of Concern – API-First imposes a formal separation of the front– and back-end. E-Enterprise involves multiple partners developing and maintaining their own front-end systems. A formal separation of front– and back-end will help ensure that the services being created are interoperable with diverse partner systems.
* Modularity – Separation of the front– and back-end frees the front-end development to combine and consume APIs and data from multiple sources.

### Common Enough Information Model

The diversity of E-Enterprise partners that can contribute to and benefit from this effort ensures that partners cannot be expected to implement identical sets of business rules, technical approaches, and data models Partners will require a level of autonomy that precludes a solution that enforces or requires an overly specific or rigid framework of its partners. Facility IPT discussions and analysis from detailed discovery sessions indicate that this effort should instead strive to define a ‘common enough’ set of business rules and data standards to make the vision of this effort and of E-Enterprise a reality.

### Some Shared Business Rules

Similar to information models, E-Enterprise partners will continue to require a level of autonomy such that partners maintain and enforce their own business rules. Some level of common or shared business rules will be essential in establishing trust and relationships between partners. For example, shared rules around primacy determination and data stewardship may be required to fulfill this shared facility data management vision.

### Sharing Good Practices and Tools

E-Enterprise creates an environment where partners can pool resources, tools, and approaches to tackle common problems. This can provide efficiencies to all partners and establishes partnerships where capabilities may need to be developed only once (as opposed to each partner independently developing the same or similar capabilities in their own environment). This benefit can be extended to software tools, shared lessons learned, and common business practices. Focusing on APIs and other reusable components that can be leveraged by multiple partners will be fundamental in the success of this effort moving forward.

## E-Enterprise Facility Governance

In addition to the information model and API technologies used to share facility data, a range of business rules must be ascertained. These rules will guide when and how data changes are negotiated and established. While some of these rules will consider lower-level issues (such as ownership of individual fields), other business rules will be high-level and concern the kinds of service agreements that EPA and SLTs must establish to formally engage in facility data integration. Because such agreements are expected to require modest business process changes for partners, approval from senior managers may be required. In this context, higher-level enabling agreements are what is meant by the term ‘governance.’

Another way to conceptualize governance is to consider the non-technical part of the Memorandum of Understanding (MOU) that would be used when an SLT and EPA wish to formally enter into routine automated sharing of facility data. The importance of partnership within E-Enterprise was identified explicitly by the State-EPA E-Enterprise Working Group’s Blueprint Team as Design and Operating Principle #1 (Partnership of Environmental Government Regulators).

Facility data integration will require governance around a specific set of rules, policies, and technical decisions. As a subset of E-Enterprise governance responsibility, facility data integration governance will fall under the umbrella of the Management Board. As other IPTs progress, this issue of common governance must enable interoperability. Governance needs associated with this effort will be identified more fully in Phase II.

# Summary and Recommendations for Phase II

This Discovery and Analysis document is the initial step in establishing services that will allow facility data to be shared across E-Enterprise. The next steps will further define the use cases, governance areas, and approaches for E-Enterprise facility-related services. The long-term success of facility services for E-Enterprise will require a solution(s) design that is fully vetted and collaboratively selected through subsequent phases of work by E-Enterprise partners.

## Summary of Findings

This Discovery and Analysis document is the culmination of Phase I of this IPT and documents the hard work and collaboration of all members. Membership in the IPT consisted of representation from twenty states and several EPA offices. Members were able to share a great deal of information and achieve a very high degree of consensus on several complex concepts and goals, which are captured in the guiding principles presented in Section 4.2.

## Phase II Recommended Approach

The IPT participants recommend a continued dialog between E-Enterprise partners. There are many concepts to further discuss, requirements to gather, business cases to refine, and prototypes to test in the next phases of this effort. Subsequent phases will continue to evaluate concepts and develop consensus.

The strong consensus behind the guiding principles provides a strong platform to build from. Next steps in this effort will seek to follow and further refine these guiding principles. As this effort moves forward into its subsequent phases, the approach will seek to be more Agile in approach and to take small steps learning and making adjustments as it moves towards its goals.

## Phase II Recommended Actions and Goals

The following sections identify the potential next steps in the establishment of E-Enterprise facility services.

### Evaluate Requirements from Related E-Enterprise Teams

E-Enterprise is comprised of several goals and initiatives. To meet the varied goals, several E-Enterprise IPTs and scoping teams have been chartered or charged to collaboratively move various aspects of E-Enterprise forward. Facility is a concept that is relevant to many aspects of E-Enterprise. Continuing efforts related to E-Enterprise facility services should evaluate requirements related to facility being established in related E-Enterprise teams. The Combined Air Emissions Reporting (CAER) project is one of several examples of E-Enterprise teams with strong dependence on the efforts of this IPT.

### Define the Facility Profile and Data Model Mapping

Phase I built consensus that a facility ‘profile’ needs to be defined in terms of E-Enterprise, and also that if partners have common enough data models that sharing such a profile is achievable. Phase II will include further analysis on which data needs to be included in this ‘profile’ and more explicitly compare partner data models to prove that they can be mapped.

### Perform Additional Detailed Discovery Sessions

Workgroup members of the first phase of this effort were very generous in offering their time to engage in detailed discovery sessions. The goal of the discovery sessions was to help improve understanding of the current state of facility management among partners, and to provide thought leadership in the ongoing efforts to define facility services for E-Enterprise. Time and other constraints limited the number of Phase I discovery sessions to three state partners. As subsequent phases are initiated, the team should consider whether performing additional discovery sessions would add significant value.

### Define Shared Service Opportunities

The efforts of the first phase of the Facility IPT have begun to surface likely candidates for facility shared services that would help E-Enterprise realize its goals and vision. The next phase associated with this effort should clearly define and prioritize specific services aligned with an API-First approach that support the partnership vision outlined in Section 4.3.1.

### Identify Opportunities and Develop Pilot/Proof-of-Concept Services

Subsequent phases of this effort will seek to identify high-priority shared service opportunities. The workgroup will also investigate opportunities to develop pilots or proofs-of-concept to demonstrate that successful shared facility services can be created and utilized by E-Enterprise partners. Prototypes may include:

* Demonstrating service-based sharing of facility profile information between EPA and a state (or states) that are providing master data management.
* Working with a state, tribe, or territory that is not currently providing master data management to prototype using FRS/CDX API services for direct reporters.

### Expand Upon Joint Governance

The governance structure identified in this Discovery and Analysis document identifies areas of governance that should be formalized in order to build a network of trust among partners. In order for this collaborative effort to successfully meet its goals for subsequent phases, this governance structure must be subject to further definition and continuous improvement.

### Develop Final Recommendations and Implementation Plan

Drawing from all performed IPT activities, including shared service definition and prioritization as well as results from pilot or proof-of-concept efforts, the IPT will develop final recommendations and an implementation plan for E-Enterprise shared facility services.

# References

* E-Enterprise for the Environment Conceptual Blueprint: Principles and Components

<http://www.exchangenetwork.net/ee/EEnterprise_Conceptual_Blueprint_013114.pdf>

* Environmental Information Exchange Network E-Enterprise for the Environment

<http://www.exchangenetwork.net/e-enterprise/>

* EPA E-Enterprise for the Environment

<http://www2.epa.gov/e-enterprise>

* EPA Leadership of E-Enterprise for the Environment

[http://www2.epa.gov/e-enterprise/leadership-e-enterprise-environment#Joint Governance](http://www2.epa.gov/e-enterprise/leadership-e-enterprise-environment%23Joint%20Governance)

* Charter for State and EPA E-Enterprise Leadership Council

<http://www.exchangenetwork.net/wp-content/uploads/2014/03/09-16-13-Signed-EELC-Charter.pdf>

* Facility Registry Service (FRS)

<http://www2.epa.gov/enviro/facility-registry-service-frs>

* EN2015 - 2015 Exchange Network National Meeting

<http://www.exchangenetwork.net/meetings-and-events/en2015/>

Appendix A. Charter for an Integrated Project Team to Scope Options for Integrating Facility Identification Data

**Charter for an Integrated Project Team to**

**Scope Options for Integrating Facility Identification Data**

*This Charter defines the objectives, planned work, and membership of an Integrated Project Team (IPT) that will provide stakeholder input into an approach for how EPA, states and tribes can share facility identifiers, correct facility data as it is being reported and facilitate many of the features and functionality envisioned by E-Enterprise.*

1. **Scoping Facility Integration for E-Enterprise**

EPA, states, and tribes are working to implement E-Enterprise for the Environment—a transformative initiative that will improve the business of environmental protection in the United States. Integrated facility identity information is critical to achieving that vision. EPA and each state environmental agency—as well as each program office (such as air, water, and waste) within those agencies— separately collect, record, and maintain both the specific data needed for each set of regulations and the core information to identify each regulated facility by name, address, geographic location, owner, etc.

As a result, facilities have to report and update the same basic identification data to multiple programs, at multiple levels of government, and at multiple times. Oftentimes, there are differences and discrepancies among these disparate sets of facility identification data that prevent accurate correlation across programs and agencies.

Environmental agencies have been independently working to better integrate their facility information. EPA uses the Facility Registry System (FRS) to gather, link, and correct facility data from individual program systems at EPA. Many state agencies have invested in their own similar systems for integrating facility identification information at the state level.

Despite improvements in methods to share, link, and correct facility information through back-end reconciliation, data on regulated facilities remains highly fragmented and incomplete. E-Enterprise offers an opportunity to explore possible approaches to integrating facility identification data across programs and agencies and correcting facility data as it is being reported. A more comprehensive approach or approaches can:

* Save time, staff, and money for both agency regulators and regulated facilities by allowing them to streamline data collection and reporting requirements and to assemble more quickly the multi-media environmental data they need for consolidated reports, permits, and inspections.
* Provide the public a more complete understanding of all the regulatory obligations and environmental impacts at each facility across media programs.
* Help agencies to manage their responsibilities more efficiently by identifying the most serious risks, setting priorities, and establishing performance measures.
* Increase data accuracy and thereby reduce the risk of disseminating incorrect information about a regulated entity.

Under the auspices of the Exchange Network Leadership Council (ENLC), this IPT will bring together partners from EPA, states, and tribes. The IPT will conduct scoping for a system of shared facility identifiers. The scoping process will document current partner capabilities on Facility data integration, define our collective future state goals and requirements and propose and test solution alternatives, with the goal of recommending potential approaches for development and testing.

**Objectives and Anticipated Work**

This IPT will explore, identify, and test possible approaches for integrating or sharing facility information across programs and agencies and improving facility data quality by accommodating data correction as it is being reported to environmental regulators. The IPT will develop an understanding of current EPA, state, and tribal efforts to integrate facility data, identify a shared set of goals for supporting more comprehensive facility integration, and identify and test options for achieving those shared goals. The IPT will not evaluate ways to improve the quality of corporate parent information. In its approach to the new data management model for a master facility identification, EPA will be developing the necessary linkages for corporate parent information to fit with the solutions identified by this IPT.

The IPT will approach its work in IV phases.

***Phase I – Discovery and Analysis***

The first phase will focus on Discovery and Analysis. The IPT will gather EPA, state, and tribal input and requirements to understand current business and management processes, data models, technical approaches, lessons learned and challenges around facility data management.

In Phase I the IPT will:

* Gather information about a diverse set of partner experiences related to facility master data management. The IPT will catalogue existing partner systems and summarize their features and characteristics. The IPT will also gather information about lessons learned on system implementation, business process changes, and governance.
* Conduct detailed 'one on one' discovery sessions with 2-3 state/tribal members to allow for a deeper dive into business processes, data models, technical approaches, lessons learned and challenges.
* Compare state/tribal business rules with existing EPA FRS services and other EPA program business rules.
* Author, review and revise an EPA/State/Tribal Facility Data Integration Discovery Document. This document will include recommendations for next steps and possible pilot options that can be used to inform follow-on phases of the IPT.

The IPT’s research and requirements gathering are critical steps in the development of E-Enterprise and reflect the commitment of the E-Enterprise Leadership Council (EELC) to work collaboratively and emphasize burden reduction through facility integration in the design and operation of E-Enterprise infrastructure. The IPT will document its findings in a set of deliverables that will include:

* Meeting minutes and summary of EPA/State/Tribal Discovery Sessions.
* Draft EPA/State/Tribal Facility Data Integration Discovery Document.
* Final EPA/State/Tribal Facility Data Integration Discovery Document.
* Functional specifications for sample shared services for State/Tribal Facility Data Integration capabilities that may be used for small pilot work in subsequent phases.

***Phase II – Define Future State Goals and Requirements***

Based on the information gathered in Phase I, the IPT will define the goals and requirements for a future state or states that will provide for more comprehensive integration or linking of master facility data across agencies. Using a LEAN process, the group will produce a future state visioning document that:

* Articulates goals for an approach to integrating Facility data.
  + What do we want to be able to do in the future?
  + What types of functionality and requirements must the future state fulfill?
* Evaluates requirements identified by E-Enterprise scoping teams.
* Defines Shared Service Opportunities.
* Considers needs for joint governance of master facility records.

***Phase III – Identify Solution Alternatives and Testing Opportunities***

The IPT will develop and document possible solution alternatives based on the goals and requirements identified in Phase II. The alternatives should include proposed technical approaches as well as relevant policy considerations. The IPT will also identify opportunities to design and execute testing with a limited number of IPT members to demonstrate the viability of possible approaches.

***Phase IV – Develop Final Recommendations and Implementation Plan***

Drawing on information and lessons learned from solution testing and pilots, the IPT will develop a set of recommendations and a plan for further evaluating and testing a recommended solution or solutions. The plan should include cost estimates, an anticipated timeline, and as many early wins as possible to show incremental progress toward the goals. As part of the plan, the IPT should present a Return on Investment analysis that estimates the qualitative and quantitative costs and benefits of proceeding.

1. **IPT Structure and Membership**

IPT members will be recruited from EPA, states and tribes. Members should possess a strong background and knowledge of facility data and their usage across environmental programs and agencies. During the course of the work, IPT members may be asked to invite other subject matter experts from their organizations to enrich the conversation. These subject matter experts may include individuals with experience in particular environmental programs or business processes, policy formulation, Information Technology, software engineering and design, administration, etc.

Lee Kyle of EPA’s Office of Environmental Information will be EPA Co-Chair and Kim Hoke of the Missouri Department of Natural Resources will be state Co-Chair of the IPT.

IPT meetings will focus on the objectives listed in this document. The IPT will work with an EPA technical contractor(s) with expertise in facility master data management approaches. EPA will task the contractor(s) with creating deliverables and eliciting input and feedback from the IPT.

The IPT will report directly to the ENLC and the co-chairs will update the ENLC and the E-Enterprise Leadership Council (EELC) on the IPT’s activities.

1. **Roles and Responsibilities**

**IPT Members:** Members will actively provide input during IPT meetings, and review and comment on deliverables. Members should make every effort to attend every call or, if unavailable, arrange for participation by an alternate. IPT members will discuss deliverables within their organizations for topics that fall outside their area of expertise.

**Co-Chairs:** With support from the IPT Facilitator and the Technical Contractor(s), the co-chairs will be responsible for communicating with other members and will lead discussion in calls. The IPT Co-Chairs, in consultation with the members, will determine the meeting process, schedule, and topics. The Co-Chairs will also be responsible for making decisions or resolving issues as necessary to keep the IPT focused on its objectives. The IPT Co-Chairs will work with the ENLC co-chairs to report progress to the ENLC.

**Technical Contractor(s):** At the direction of the EPA Contract Officer Representative (COR), the technical contractor(s) will be responsible for leading the IPT’s technical discussions and eliciting and documenting input from the members. The technical contractor(s) report directly to EPA’s Office of Environmental Information and will act as Facilitator for the meetings including managing day-to-day operations and logistics for the IPT including production of agendas and meeting summaries, call scheduling, and other support tasks.

**EPA Contract Officer Representative:** EPA Contract Officer Representatives (COR) or their designees, in consultation with the IPT co-chairs, are responsible for providing all formal direction and tasking to the technical contractor(s).

**Exchange Network Coordinator:** The EN Coordinator will serve as an additional staff resource for the IPT and work closely with the Co-Chairs to communicate with the ENLC.

1. **Meeting Frequency and Schedule**

The IPT will meet at least every 3 weeks. That frequency may be adjusted at the discretion of the Co-Chairs. Given the diverse location of the members, the IPT will predominantly meet via conference calls.

The IPT will work with EPA and its technical contractor(s) to establish a schedule for the work and report back to the ENLC. The IPT Co-chairs and the co-chairs of the ENLC should discuss any possible changes to the schedule.

Appendix B. Use Cases

From the viewpoint of a regulated entity, the following first touch topics were reviewed:

* First touch/initial contact process depends on the entity type.
* The general scenario that generates the first touch may be a permit. This could include an air Point Source Pollution (PSP) permit or a storm water construction permit at the state level.
  + The permit may be proactively sought by the industry, obtained via a referral from another department/agency, or the result of enforcement by a co-regulator.
* Who the regulated entity representative goes to depends on the environmental laws/delegation.
* The numbers and types of permits can vary significantly from state to state.
  + States with more industrial activity are more likely to have large campuses with complex inter-relationships among different media/programs. Other states may see that a particular permit to small local governments dominates their statewide permitting landscape.
* For large projects, the Environmental Impact Assessments (EIAs) process usually determines which regulatory agencies will be involved (i.e. EPA vs. state, or both).
* Facilities in fully delegated states do not typically interact with EPA, unless there are compliance issues where EPA has enforcement authority (ex. cease and desist for air/water/waste that requires a site to apply for a permit).
* Facilities will have unique needs with respect to facility and cross-media integration.
  + Campus-type facilities often require a sophisticated reporting paradigm.
  + Airports usually comply with many different regulatory requirements.
  + Military sites (ex. army or navy) are federal facilities that often involve coordination with both the state and EPA (ex. Army Corp Engineers may be involved).
* For certain EPA direct reporters (ex. Toxics Release Inventory/TRI), there may be regulated entities interacting with multiple agencies, none of which coordinate at this time.
* Compliance is not only dependent on the site type (ex. airport vs. direct reporter to EPA vs. small business requiring a single state permit) but also business type (ex. a single company may report to multiple state agencies because it does business cross-state).

Potential regulated entity use cases were identified:

* *New facility* without pre-existing information in EPA/state systems either needs to perform an action (ex. obtain a permit) or wants to know what to comply with (i.e. compliance assistance request). The facility may not have a street address yet, only geospatial coordinates (ex. plot of land or site).
  + Consideration: Is differentiation needed if the ‘first touch’ is the result of enforcement or compliance assistance?
* *False start* where a process that might require a record is started in a co-regulator’s system but is later determined unwarranted (ex. there is nothing to regulate or no impact to report).
* *Pre-regulator* information is generated by a municipality/others (ex. a resident may request a permit to build a deck, and in the process discovers that a wetlands permit is required as well).
* Facility (ex. port) that reports to *multiple agencies* which do not coordinate. The pain points for such facilities are expected to be different from pain points at facilities that report to a single agency.
  + Consideration: In a future phase, potentially interview stakeholders and document their problem statements to glean insight. Pain points are expected to vary for facilities that report to a single agency vs. multiple agencies. In general, it is difficult to drive users to the same portal and then to disperse these consumers to different agencies.

Similarly, from the state/EPA perspective, the following use cases were identified:

*A record exists somewhere*. A record exists somewhere within the agency or within the enterprise but external to the agency and E-Enterprise partner notification of modified data may be beneficial.

* Determine how to account for the *varying levels of information based on need*.
* *Public notice* information that leads to a request for different kinds of information.
* Users typically want to search by either an address/geographic polygon location or by a ‘thing’ (ex. Facility).

Appendix C. Discovery Session Introductory Letter

Facility Integration Integrated Project Team Co-Chairs

October 26, 2015

Discovery Session Participant

New Jersey Department of Environmental Protection, Wyoming Department of Environmental Quality, Missouri Department of Natural Resources

Dear Discovery Session Participant:

On behalf of everyone participating in the Facility Integration Integrated Project Team (IPT), thank you for agreeing to participate in detailed discovery sessions, or ‘deep dives’. The information gathered during these discovery sessions will be critical to the discovery and analysis phase of this effort. Recognizing that each participating organization is at a different level of maturity in their facility management approach, we would like to request that you provide any of the following materials (related to the overall management of facilities and their information) that are readily available at your earliest convenience and prior to your scheduled discovery session if possible:

* Process documentation for inbound collection/correction, processing, integration, management, and etc.
* Standard operating procedures for the above processes.
* Data governance procedures/processes for defining and ongoing coordination of cross-program facility data issues.
* Logical data model or entity relationship diagrams.
* Physical data models.
* Terminology dictionary and/or data element definitions.
* Best Practices for any of the above.

This request for information is not a request for your organization to create any materials, but rather to provide for analysis any existing materials that may be helpful to the IPT during this discovery and analysis phase. Please email the information/documents to:

* Kimberly Hoke ([kimberly.hoke@dnr.mo.gov](mailto:kimberly.hoke@dnr.mo.gov))
* Lee Kyle ([Kyle.Lee@epa.gov](mailto:Kyle.Lee@epa.gov))
* Luke Gentry ([lucas.gentry@cgifederal.com](mailto:lucas.gentry@cgifederal.com))
* Jimmy Mahady ([jmahady@rossstrategic.com](mailto:jmahady@rossstrategic.com))

The IPT Co-Chairs will be contacting you soon in order to schedule the meeting times and participants for your discovery session. Thank you for your willingness to participate in this important discussion.

Sincerely,

Facility Integration Integrated Project Team Co-Chairs

Kimberly Hoke, Missouri DNR

Lee Kyle, US EPA

Appendix D. Discovery Session Agenda

Master Facility Data Management Discovery Session

[Organization Name]

[Month DD, YYYY]

Meeting Preparation

Organization to provide any materials, if they exist, prior to the meeting including but not limited to:

* Process Documentation
* Standard Operating Procedures
* Data Governance Procedure
* Logical Data Model
* Physical Data Model
* Terminology Dictionary
* Best Practices

Please email the information/documents to:

* Kimberly Hoke ([kimberly.hoke@dnr.mo.gov](mailto:kimberly.hoke@dnr.mo.gov))
* Lee Kyle ([Kyle.Lee@epa.gov](mailto:Kyle.Lee@epa.gov))
* Luke Gentry ([lucas.gentry@cgifederal.com](mailto:lucas.gentry@cgifederal.com))
* Jimmy Mahady ([jmahady@rossstrategic.com](mailto:jmahady@rossstrategic.com))

Meeting 1

| **Topic** |
| --- |
| Introductions   1. Could each participant from your state provide a brief introduction – role, number of years with the agency, and years of experience with master facility data management?   Level of Experience Overview   1. How long has your agency been doing master data management (MDM)? 2. What was the original goal or pain point that led your state to begin to integrate your data? What problem were you or are you trying to solve? 3. What is a brief history of your work with master data management? |
| Organization Demonstration   1. If your organization feels it would be useful to demonstrate systems or processes related to MDM, please let the Co-Chairs know before your scheduled discover session. |
| Business Processes (Focused based on materials provided if possible)   1. Can you provide an overview of the facility data entity universe?    1. Please define how you interpret the terms ‘facility’ and ‘thing of interest’? For example, do you sum things at an abstract level such as building or lot?    2. How do you assign IDs? At an abstract ‘facility’ level? Or also at a more detailed level such as something like a boiler within a building?    3. How many ‘facilities’ and/or ‘things of interest’ do you regulate? 2. Who are your main stakeholders/customers for master facility data? Internal staff, external regulated entities, the general public, a combination, etc.? 3. When do you perform data reconciliation and correction? (Ex. At the intake? Reconciled at end of process? Some hybrid?) 4. When, if ever, is master data presented to users in your systems? Do you enable reporters to pre-populate forms and correct data as it is being reported? Does your system help reporters enter high-quality data? 5. How do you handle Facility ID assignment (master and program-specific)? 6. Do you currently share facility data with EPA or other agencies? 7. What are your facility data management staffing approaches? Do you have any dedicated staff? 8. What formal processes, operating procedures, and best practices do you utilize related to reconciliation and sharing/integration of facility data? 9. What governance process do you use to determine how data is shared, integrated, and managed? 10. Do you have a retention schedule? Is there ever a point at which you no longer ‘count’ these regulated entities and how are those data managed? 11. How do you manage a facility that closes or where key information, such as owner, changes? 12. How often do you attempt to compare state data with federal facility data? |
| Data Model (Focused based on materials provided if possible.)   1. Logical data model    1. What core entities and types of data do you manage with master data management?    2. What level of relationships between core entities are supported?    3. What terms do you use when referring to core or master data?       1. What is the level of consistency with respect to terms used by other agencies in your state?       2. How well are those terms used across your agency? In other words, has your agency adopted this way of thinking about and talking about your entities, or have they kept their own ‘language’?       3. How are term dictionaries/definitions accessed? 2. Physical data model (The physical data model, if provided, will likely be used primarily for offline analysis or as a supplement to the logical data model discussion)    1. What are the known limitations or challenges based on the data model? |
| Technical Approaches   1. What 3rd party data management solutions do you utilize? 2. To what extent do you provide web services or other data integration tools related to master data management? 3. Are there publically available resources to access reported facility data? |
| Lessons Learned and Challenges   1. What metrics have you used to demonstrate that the objectives of the program are being met? 2. Based on your experience to date, what have you learned?    1. Topics may be related to:       1. Data and data model       2. Regulated entity behavior       3. Management/collaboration       4. Staffing       5. Regulatory opportunities and barriers 3. What challenges have you faced, and how did you overcome them? 4. What difficulties do you continue encountering? What solutions have you developed to resolve these challenges? 5. What do you think are the biggest hurdles to more fully integrating this data across the enterprise? |

Meeting 2

|  |
| --- |
| **Topic** |
| Meeting 1 Recap   1. Summary of Meeting 1 topics discussed. 2. Sharing of any initial analysis or response. |
| Organization Needs/Perspective   1. Open agenda item for participating organization to share any requirements, needs, questions, or other feedback on the discovery sessions or IPT as a whole. |
| Follow-up Discussion based on Meeting 1 discussion and materials |
| Wrap-up |

Appendix E. Burden Reduction for Direct Reporters – the CDX/FRS Facility Management Widget and TRI

The Toxics Release Inventory (TRI) has worked with both the Central Data Exchange (CDX) and the Facility Registry Service (FRS) to develop a widget embedded within CDX that is used to manage facility information. The widget allows direct reporters to search for facilities within FRS and associate it with their CDX profile. Information about the facility can also be managed within the widget. Data elements such as facility name, address, and location can be changed within the widget and updates sent to both TRI and FRS. The development of the widget has led to dramatic time savings for both direct reporters and EPA for all programs who have adopted use of the widget to support direct reporting needs.

In the case of TRI, direct reporters who leverage the FRS API save about **140,269** **hours total for a reporting year**. In Reporting Year 2014, a total of approximately 23,000 facilities reported to TRI. The total hours estimate is based on three cases for a facility:

1. Facilities doing revisions – The TRI program estimates that for a given reporting year, 1,500 facilities do revisions to TRI submissions. These facilities will save about 25% of the estimated burden since they will not have to re-enter facility data accessed through the widget. The calculated burden for Form R is just over 120 hours, so that would amount to a savings of 30 hours per facility. The total savings for this subset of facilities would be **45,175 hours**.
2. Non-reporters providing facility level information without filing a form – The TRI program estimates that 600 facilities spend on average 3 hours a year doing this. The total savings for this subset of facilities would be **1,800 hours**.
3. Facilities that use locational API services in the widget – The TRI program reported that in RY 2014, 7,750 facilities used the locational API service and about 10% of their time is generally spent verifying and correcting data without locational API services. The total savings for this subset of facilities would be **93,294 hours**.