E-ENTERPRISE for the environment

Modernizing the business of environmental protection

Business Rules to Support Integration of Facility Information 1.0

May 2018

The Facility Business Rules group, a work stream of the E-Enterprise Facility Integration that began in Phase II, documented business rules to enable integration of facility information across programs, among partners, and across the enterprise. The team consisted of members from 11 States and several U.S. EPA programs (see list of participants in appendix). The members provided specific information on how their programs use facility information and reflected on the business rules necessary to integrate facility information across various existing systems. These business rules will continue to be refined as pilots are conducted and facility integration shared services are developed. The group looked at the following areas in particular:

- *Clean air (including Emissions Inventory System (EIS), Toxics Release Inventory (TRI), Greenhouse Gas Reporting Program (GHGRP), Integrated Compliance Information System for Air (ICIS-Air), and South Carolina and Oklahoma programs),*
- *Drinking water (New Hampshire program),*
- *The National Pollutant Discharge Elimination System (NPDES) (including Colorado, Alabama, Iowa programs), and*
- *EPA's Facility Registry Service (FRS).*

This document contains descriptions for the following business rules:

Other supporting documents, including individual inputs, can be found on the team's SharePoint collaboration space (access for E-Enterprise members and project teams).

Introduction

The Business Rules group created the facility business rules described in this document to be broadly applicable and support E-Enterprise facility information integration across different partner applications. Going forward, these business rules can be used as a reference for and tested by specific projects like Combined Air Emissions Reporting (CAER) or future Facility pilots. The business rules are designed to accommodate all programs involved in facility integration and thus, need not be biased or favor one program or set of programs over another. Project testing will provide more evidence of the applicability of the business rules under various scenarios and could also reveal that some business rules might need to be revisited if potential downstream or integration conflicts are identified. The business rules will also be further refined based on work done in subsequent Facility Team work streams and pilot projects.

The business rules contained in this document should be viewed at a conceptual level. This document does not contain a complete, comprehensive, and detailed set of business rules for facility integration. Application of these business rules to specific business processes may require additional thought and lead to further refinement of the business rules. It is also not possible to articulate a comprehensive list of detailed business rules at this time, given the highly technical information that would need to be incorporated into them. Similarly, some of the recommended business rules might also depend on data availability (e.g., those related to data completeness). States and Tribes can integrate locally with current available data, as applicable. Shared services that support integration should be consistent with the E-Enterprise Shared Services Strategy *(SharePoint access for E-Enterprise members and project teams)*.

Members of the Business Rules work group have arrived at consensus on these business rules. The Business Rules workgroup realizes, however, that such a document may not carry enough information or meaning for more experienced or technical users. Because of this, a goal moving forward is to include appendices for each business rule where greater explanation or detail is provided to fully comprehend the nuances of the topic, and to properly reflect relevant technical aspects of the rules.

One of the most difficult tasks in developing these business rules was to ensure that participants were discussing the same topic from the same perspective at the same time when such a large and varied set of systems, data contributors, and integration uses are encompassed by facility integration. While we suggest reviewing this document from individual programmatic or use case perspectives, readers must also view these business rules from the perspective of the general goal of *facility integration*. Because the scope here is rules that will support individual needs ***and*** enable facility integration that is useful to all, it may appear that these needs conflict at times.

The functions of the FRS, program systems, and state systems used in this analysis were the "as-is" functions that were in production in the summer of 2017. As these various systems are enhanced over time, modifications may need to be made to these business rules. By making these rules broadly available now, it is the intent of the Facility Team to influence any planned system enhancements by E-Enterprise partners to align with this common set of business rules for facility integration.

The Facility Team's Phase III will further assess the benefits and challenges of approaches to facility data integration. Approaches to be reviewed will include:

1. **Facility integration through shared services centrally managed to support State, Local, and Tribal (SLT) partners who do not have facility integration infrastructure**
    - Example: State, Local, and Tribal (SLT) partners who do not have facility integration systems or business processes in place and opt to leverage shared services and a centrally managed system, such as the EPA's FRS.

2. **Facility integration through shared services centrally managed to support SLT partners who do have facility integration infrastructure and where both program facility records and partner integration records are mirrored but subordinate to a centrally managed, integrated facility record**
    - Example: SLT partners that integrate facility data through a general state/local service and not through an environmental program perspective (e.g., the SLT combines environmental facility information with other services such as tax, licensing, etc.), which makes SLT integrated records diverge from records coordinated in a centrally managed integration system.

3. **Facility integration through shared services centrally managed to support SLT partners who do have facility integration infrastructure, where both program records and partner integration records are mirrored in the centrally managed service, and where the partner's integration record is operationally equivalent to a centrally managed, integrated facility record**
    - Example: If SLTs have facility integration infrastructure in place, the same data in should yield the exact same data integration out. Depending on the system in place and logic employed, SLTs that may initially be identified as this scenario might, after analysis, fall under the second or fourth scenario.

4. **Facility integration through a distributed network of partners implementing shared services where an index (or indexes) is (are) mirrored across the partner systems (referencing both program records and a partner's integration records).**
    - Example: Distributed integration through various shared services, common goals, and business rules where all participating parties proactively engage in facility integration.

The first two scenarios represent a traditional vertically integrated business model, but at this point *these proposed high-level business rules do not make that assumption (e.g., integration using U.S. EPA's FRS)*. Depending on the facility data integration approach, business rules A (Facility Record Definition) and B (Data Ownership, Data Stewardship, and Collaborative Use) laid out in this document will need to be adjusted accordingly.

A. Facility Record Definition

The definition of facility should stay within the purview of the originating environmental program.

A1. The originating environmental program defines the required content of their facility record.

A1.1. States and Tribes may perform facility integration within their own systems.

A1.2. Originating environmental program refers to the primary data owner that manages the environmental program (see below under business rule B for definition of the primary data owner). An originating environmental program may overlap with another environmental program system that includes data systems (e.g. CEDRI, EIS, TriMeWeb) that are part of other environmental programs and thus house data that should/should not be integrated (i.e., from a state that already has facility MDM).

A1.3. In these business rules, **"Facility"** refers to the environmental program facility record, not to a physical facility.

A1.4. The environmental program facility record is the record of the facility that represents the environmental interests of that facility for that program and is created by the originating environmental program (i.e. originates from an environmental interest).

A2. For record creation and operational use, each environmental program facility record must meet the definition of facility within its originating program and include information for the mandatory data elements defined below under Business Rule C.

A2.1. A facility is in operational use if its permit status is in operation, which indicates if facility records meet certain thresholds prior to being used in integration algorithms.

A2.2. An **integrated facility record** can represent the current record or the entire record lineage. S/L/T or EPA programs with an integrated facility record might have multiple environmental program facility records associated with it.

A2.3. An environmental program should provide the definition of facility within that originating program.

A2.4. To the extent possible, optional data elements listed in the Business Rule C should also be included.

B. Data Ownership, Data Stewardship, and Collaborative Use

Clear communication about authorship and ownership is required. For ease of application duplicate datasets may exist as long as appropriate citation is used. This will prevent change upon change and updates being made because of differences in how each program sees a facility.

B1. For E-Enterprise Facility Integration to succeed and support the various uses envisioned, respectful use of data with topical deferral to the source that data originated from is required. This implies that for every piece of data and metadata (as opposed to sets of data or metadata), the author of that data must be known and clearly communicated to all parties in all situations for all uses.

B1.1. This requires knowing explicitly the domain and scope of each environmental program and the interplay between various environmental programs so that proper citation may be given.

B1.2. An author is any individual/agent/entity that produces data. Generally, authors submit data to at least one environmental program. Authorship is an independent action, that is, an author maintains complete control of the form and content of the final, published, data product being authored. An author may exist outside of or within (as staff) an environmental program.

B1.3. The **primary data owner** is the individual/agent/entity who is charged with responsibility for environmental regulations.

B1.4. An editor is any individual/agent/entity that amends/edits/modifies existing data and/or the meaning of existing data. This may occur with or without authorization from the primary data owner. Note that an author is responsible for making corrections or amendments to data in a timely manner, if an error is reported by a data steward or user. In this context, after an author has released control of the data product, an author transitions to editor.

B1.5. The FRS (or other integration system/service) is not meant to be the system of record for environmental program facility records, but FRS services will help establish relationships between environmental program systems and environmental program facility records more timely and efficiently.

B2. In cases where it might not be practical for data users to directly tie to a primary source, duplication of primary source data with appropriate citation is required, and that duplicated data may be subsequently tied to or referenced by others.

B2.1. Cases where it might not be practical for data users to directly tie to a primary source include, for example, instances of obtaining a dataset to maintain a specific temporal, spatial, or relational characterization of that dataset for complex or resource intensive analyses. Because the obtained dataset may quickly lose congruence with source material, citation, versioning, and relation to source material should be explained clearly and concisely.

B2.2. Duplication, subsequent editing, and business processes normally associated with program management can produce datasets that contain primary source data, secondary source data, and tertiary source data. To avoid any confusion among data users, citation should be resolved to individual data points within these datasets.

C. Data Completeness: Mandatory and Optional Data Elements

Mandatory and optional data elements are needed to ensure consistency and enable data linkages and reconciliation across programs. States and Tribes can integrate locally with current available data, as applicable.

C1. Facility records must contain the below mandatory fields when submitting a facility to the shared enterprise. These fields are considered mandatory from a facility integration perspective.

- **Facility name**: the full facility name at the physical location should be included. For integration purposes, the FRS facility record can be referenced. A facility name is the public or commercial name of a facility site (i.e., the full name that commonly appears on invoices, signs, or other business documents, or as assigned by the state when the name is ambiguous).
- **Program name** (e.g., EIS, NPDES, Wetlands, Onsite Wastewater Treatment Systems (OWTS), etc.)

- **Program ID** (The identification number, such as the permit number, assigned by an information management system that represents a facility site, waste site, operable unit, or other feature tracked by that Environmental Information System).
- **Location coordinates** should include a facility's address location and/or latitude/longitude, and/or township, section, and range.
  - **The address location** is the address that describes the physical (geographic) location of the front door or main entrance of a facility site, including urban-style street address or rural address. Address location should include the street, building number, city, state, the ZIP code, and the Federal Information Processing Standards (FIPS) county code. FIPS is the code that represents the county or county equivalent and the state or state equivalent of the United States. Standardization of addresses may use services such as Navteq and USPS data, which will allow for a standardized data before submittal. Third party implementation of USPS is also valid. In the lack of any standardization criteria, facilities can use the EPA Data Standards as the starting point. Likely to apply are: Contact Information; Facility Site Identification.
  - **Latitude/longitude** in decimal degrees. If the latitude/longitude doesn't represent the centroid, the facilities could provide the reference code (e.g., the FIPS county code) and/or other available information (e.g., section, which represents an approximately one-square-mile block of land). Best practice is that the latitude/longitude may include the horizontal and vertical measure metadata, and the coordinate system and datum. In case there is a lack of standardized criteria, facilities can use the Latitude/Longitude EPA Data Standard.
  - **Township, section, range.**

C1.1 Additional mandatory fields may be required, depending on specificity of each program and/or goal of facility integration.

C2. Facility records should contain optional fields to streamline and simplify integration. Specific provisions for these optional fields should be viewed as guidelines, and not as mandatory requirements. The below proposed optional fields do not imply that they are optional from a program perspective (since mandatory data in one program may be optional in another), but they are considered optional from a facility integration perspective at this point.

For more details on data standardization, refer to related Business Rule E.

- **Legal description**: company owner/operator including a facility owner name, facility owner's mailing address, and contact information. Legal description is generally required by all EPA programs and State/Local/Tribal (SLT) programs.
- **Status of Facility**, as defined within the originating program
  ***Note**: This optional field is still highly unknown in the context of facility integration and it's still a new field within the FRS data model. There are different uses and different definitions of "status of facility" across programs and media. Additionally, a facility could be considered active in one program and inactive in another program, e.g. some programs may look at the facility's operation status (i.e. in operation/shutdown) and others may look at the permitting status. The Business Rules team plans to further look at the "status of facility" at a later stage once more knowledge may be available.*

- **EPA Region**
- **The North American Industry Classification System (NAICS) code or the Standard Industrial Classification code (SIC).** SIC is still being used by some States and programs, though it's expected to be discontinued in the future.
- **Day-forward historical background**. All facility changes/updates from the moment the facility record was created in the system should be available in the facility profile. This information necessarily includes start and end dates for each aspect of a facility recorded for a program.
- **FRS should maintain historical facility data for all programs** to enable automation integration and to provide historical reference for users that are referencing any facility data. The program systems would also maintain their own history in their respective system or repository (e.g. possibly CROMERR).

C3. All facility changes/updates and other historical data should be available in the facility profile and will be tracked day forward from creation of the record.

C4. Where applicable and possible, sub-facility data should be included. Depending on the context of integration use cases, sub-facility data might be mandatory.

C4.1. The context and specificity of sub-facility data varies depending on the individual program. For example, if a "facility" for a water program is an entire city's water system, then a "sub-facility" for that "facility" may be a water treatment plant. That same plant may be a "facility" in an air program that could record "sub-facility" data about a water retention pond as a unit within that facility. Military bases could be another example of complex facilities that need to have sub-facility information.

Another example could be the Risk and Technology Review (RTR) project that leverages CEDRI and the FRS widget to: a) streamline and standardize RTR collection of sub-facility attributes (a.k.a. components) via CEDRI; b) share RTR-collected facility attribute data with other programs via FRS (the data would sit in FRS and marked as a CEDRI record); and c) provide a mechanism and incentive for industry to share data. A link to the RTR documentation, once completed, will be added to this example.

The Facility team plans to do a pilot in the future to get sub-facility data into FRS via either the Facility Identification (FacID) flow or by extending the services that are being developed within the Rhode Island pilot to sub-facility elements.

C4.2. A complex facility (e.g., refinery) would include sub-facility information on air emission units and release points; tanks; permitted industrial landfill; hazardous waste treatment units; monitoring wells; wastewater treatment units; etc.

D. Metadata

Metadata for each facility record will allow facility integrators to identify the data source, assess the data quality submitted by the originating program, and enable integration. Some metadata will be entered

manually, some will be derived. In all cases, metadata used in facility integration are required elements of a program facility record.

D1. Descriptive metadata is intended for understanding program facility record elements/data source. Examples include:

- user ID. User ID could be the program system user or the facility system user.
- definition of facility within the originating program
- source of data (who created, who updated the record),
- dates when the record was changed/updated,
- GPS metadata when adding latitude and longitude (precision/source, descriptive location, e.g. "front door").
- Nationally Defined Values for Geographic Reference Points code (e.g., 020 = Facility Centroid; 021=NE corner of parcel, etc.).

D2. Administrative metadata that is intended for rendering and managing program facility records, such as: type of record, how and when the record was created, the levels of record permission (including who can access, edit, delete it). FRS will respect the level of record permission that the program set.

D3. Structural metadata that is intended for integration purposes, focused on determining linkages between program facility records. Examples include information that describes how information is derived, manipulated, and integrated (e.g. standardization, abbreviation, etc.).

D3.1 Structural metadata would be further derived based on integration needs identified by ongoing pilots as facility shared services are developed.

E. Data standardization

Programs use various conventions for their facility records. Standardization will help facilitate data integration. These business rules do not create mandatory data standard requirements for programs, but the standardization criteria used must be identified to enable integration. The integration standardization will be determined by the originating programs.

E1. Originating programs determine the standardization criteria to populate the required data elements. That means an originating program can choose the standardization criteria they want to use when reporting, as long as they declare what criteria they use. The EPA data standards can be used as reference: https://www.epa.gov/data-standards/find-epa-data-standard. Specifically, some of the following data standards that are likely to apply to the facility data are: Contact Information; Facility Site Identification; Latitude/Longitude; Permitting Information; Representation of Date and Time; SIC/NAICS; Tribal Identifier; Well Information.

E2. Originating programs must specify what standardization criteria they are using.

E3. If originating programs don't have standardization criteria, the recommended criteria in Business Rule C can be used as reference.

F. Data validation

High quality data needs to be available for integration. Validation is needed to ensure that submitted facility data meets certain QA/QC criteria.

F1. Data validations should occur as early as possible, ideally during data entry. Data with errors should ideally not be allowed in, with errors flagged for correction before submission.

F2. Prior to sharing a facility record, minimum validation against the mandatory and, where applicable, optional data elements (*listed in proposed Business Rule C*) is needed.

F3. Validation can be specific to each SLT or EPA program.

G. Creation of a New Facility Record

For integration purposes, it is important to know from where the data originates in the case of new program facility records and determine cases when connections between new and existing records should apply.

The following business rules are program-centric and there may be other business rules that need to be considered before a new program facility record is created by a program at the state or EPA level particularly when taking into account the program users (e.g. EPA/State staff performing data quality work).

G1. The originating program determines if a new facility record is warranted within their originating system.

G2. Before a new program facility record is created, the originating program should check if that record or a similar one already exists within the enterprise.

> **G2.1.** Currently, EPA's FRS has a web service to support this business rule (see the FRS Query Page).

G3. Where possible, if a related record exists, the program should reuse/reference the existing record. Related records are dependent upon integration use cases.

G4. When not possible to use an existing record for reference, create a new program facility record and indicate why the existing record wasn't used.

> **Note:** If a program facility record is de-linked, FRS will not try to link the original program facility record back to the FRS facility record it was de-linked from. If a program facility record A was de-linked from an FRS facility record and later a new program facility record B is (correctly) linked to the same FRS facility record as record A, in certain scenarios that FRS facility record (which the two program facility records fall under) could be flagged as a duplicate of another FRS facility record.

H. Modification of a Facility Record/Hierarchy Decision Modeling

To establish a clear hierarchical decision process, it is essential to identify who can edit the data, who has primacy over the program facility record, what is the relationship between primacy-overwrite-timing, and what is the link back to the primary source. The hierarchy decision modeling is still an area that needs to be further tested and might be program-specific.

H1. Editing Data Elements

H1.1 Originating programs should flag/indicate that there's been a change to a facility record. Originating programs have their own copy of a facility record and there will be one shared version of the facility record in the shared enterprise.

H1.2. Originating programs should initiate notifications for changes to a facility record (including data deletion, creation of new/modified facility, merging, matching, etc.) to data exchange or integration partners (defined by programmatic or agency need). This will trigger a notification from the data exchange to systems that reference the record, that the record was updated.

H1.3. Data is authored. Authorship may be a consultant producing and submitting a report on behalf of a company. Authorship may be a legal entity's application for obtaining a permit. Authored data may be any piece of data submitted to an environmental program within the domain and scope of the environmental program and where that data did not exist prior. Included within authored data is any data environmental program staff may produce during the course of their duties. The action of systematic management/regulation of data implies that managers/regulators author original content in their regular course of business.

H1.4. Data is amended/edited/modified. This can range from correcting a data entry error to amending a contractual agreement between parties which requires numerous changes to subordinate datasets. In every case, data must exist prior to any change, and change implies that the modified data carries at minimum different meaning. Appending metadata and clarifying existing data by adding comments are examples of data edits that should not be identified as amending/editing/modifying data.

H1.5. Data is managed/regulated. The regulatory responsibility of an environmental program requires data to be submitted to it, whether authored by individuals from outside of the program or produced from within the program by program staff, and that data be maintained by and immediately accessible to program staff. At the point of regulatory responsibility, data is usually considered to be the authoritative data (as authors have submitted their data directly to an environmental program), the data management system is usually considered the system of record (maintaining data that supports official functions or data that is legally binding), and the environmental program is usually considered the authoritative source or primary source (data here is an outcome of the environmental program, that is, without the presence of the environmental program and its regulatory responsibility, the data would not exist).

H2. System Relationships – Hierarchy Decision Modeling

H2.1. Any program that does not create its own source data should strive to maintain a real-time link to the nearest upstream process/source database/system. This will result in real-time notifications to other programs (mechanism to be established).

H2.2. Information in both upstream and downstream programs should never be overwritten by linked sources unless explicitly determined as appropriate by program management of a system of record.

I. Data Deletion/Removal/History

For historical reasons, information should not be deleted if it is tied to other records, as existing linkages should be maintained. Development of other potential business rules related to data deletion/removal/history will depend on additional work needing to be done to define the use of "status" in FRS.

I1. Originating programs' data should be archived and not be permanently deleted (this implies "soft delete").

PLACEHOLDER: Upcoming topics that need to be addressed

- Status of Facility (currently included as part of the business rule on "Data Completeness" under optional fields): This is still highly unknown in the context of facility integration and it's still a new field within the FRS data model. There are different uses and different definitions of "status of facility" across programs/media. Additionally, a facility could be considered active in one program and inactive in another program, e.g. some programs may look at the facility's operation status (i.e. in operation/shutdown) and others may look at the permitting status. The Facility Integration Team plans to develop a micro-service for a unified facility status supporting all environmental programs in the next project phase (June 2018-June 2019).
- Data governance issues will be explored within upcoming Facility workstreams and incorporated in this document after real data testing and evaluation of pilots.
- Schema integration.
- Merging/Unmerging of records that are in the shared enterprise facility system/environment.
- Downtime in cases when the enterprise facility repository is down.
- Data validation (related to business rules F1 and F2) and potential fields that might need to be added to document that validation has been done (e.g. contact person, etc.).

APPENDIX: Facility Business Rules Participants

Name	Organization
Ron Evans	EPA, Co-Chair
Joshua Kalfas	Oklahoma, Co-Chair
Lynn Barnes	South Carolina
Michael Beatty	Texas
Carla Bedenbaugh	South Carolina
Joe Carioti	EPA
James Chilton	CGI Federal
Eric Cleckler	Alabama
Courtney Cswercko	Iowa
Mary Curtis	EPA
Brittany Decker	South Carolina
Sally Dombrowski	EPA
Christopher Dunbar	New Hampshire
Laurie Fleet	Texas
Beth Graves	ECOS
Carolyn Greenough	New Hampshire
Shana Harbour	EPA
Kim Hoke	CGI Federal
Ashley Inserillo	New Hampshire
David Jacobson	EPA
Matthew Kelly	EPA
Veronica Kenkel	Colorado
Mary Kimlinger	Minnesota
Arno Laud	Maryland
Josie Lopez	EPA
Anthony McClard	South Carolina
Carol Miller	South Carolina
Christy Monk	Alabama
Kaitlin Murphy	New Hampshire
Theresa Sabbia	New Hampshire
Tobias Schroeder	EPA
Leslie Simpson	Colorado
Jill Slain	Wisconsin
Robert Ward	EPA
Mark Wert	Massachusetts
Chad Wilbanks	South Carolina
Rob Willis	Ross Strategic
Dana Stefan	Ross Strategic