Facility Identification Data Exchange

Flow Configuration Document

Version: 3.0a
Revision Date: May 28, 2013

Prepared for: ECOS

Prepared by: Windsor Solutions, Inc.

Environmental Information

eXchange Network

Table of Contents

1 Acknowledgements

This document was prepared with input and support from the following individuals:

Facility ID Integrated Project Team	
Tom Aten – WI DNR	Ken Blumberg – EPA, Region I
Dennis Burling – NE DEQ	Pat Garvey – EPA, OEI
Ana Greene – EPA, OEI	Alex Klaessig – EPA OEI
Noel Kohl – EPA, Region V	Wendy Nye – EPA, Region I
Andrew Putnam – CO DPHE	Gerald Schwandt – MN PCA
Bob Simpson – EPA, Region II	Kurt Rakouskas – ECOS
Mitch West – OR DEQ	
Contractor Support	
Bethel Helene – Lockheed Martin	Veeree Kadans – Lockheed Martin
Kevin Lyons – Windsor Solutions, Inc.	Guy Outred – Windsor Solutions, Inc.
Bill Rensmith – Windsor Solutions, Inc.	Clint Teng – Lockheed Martin
TK Conrad – Windsor Solutions, Inc.	

2 Introduction

2.1 Background

Over the last four decades EPA and States have worked to improve their integration of environmental data beginning with facility data. Many States and EPA have implemented exchanges of facility data between EPA's Facility Registry System (FRS) and similar systems in those states. Exchange Network Partners are pursuing these cross-agency and cross-system facility exchanges for the following reasons:

- To take advantage of investments States have made in integrated facility data to improve the data in EPA's FRS and vice versa through the 'cross pollination' of these systems' reconciled data.

- To provide facility data to state or EPA partners for value-added integration in their own applications and services, for example:

  - For other EPA programs, such as the Chesapeake Bay Program

  - To assess cross-jurisdictional environmental issues (e.g., to support emergency response situations that cross state borders).

  - To combine facility data across intra-state agencies (e.g., integrating pesticide storage and application facilities that originates within a state's Department of Agriculture).

2.2 How to use this FCD

This document describes the recommended approaches for the Network exchange of Facility Identification data. Based on the lessons learned from the implementation of the previous version of this flow, the state/EPA Facility Integrated Project Team (Facility IPT) revised the XML structure and data services required to support this flow.

This FCD document identifies and defines the data services that are necessary to fully support the exchange of data from a Partner's Node.

2.3 Disclaimers

This FCD describes a pure 'publishing' flow of Facility Identification data. A publishing flow allows a flow implementer to offer their data up to many (authorized) exchange Partners, and those partners 'pull' data of interest as and when they need it. While this flow provides some data services that support ad-hoc, targeted requests for select data, this flow also allows employs some 'incremental update' data services that support Partners that wish to maintain a local, synchronized replica of facility data. As all published data that is exchanged is initiated by a Query of Solicit request (and not Submit), there is no need for a payload header.

3 Component Alignment and Change History

3.1 Flow Component Version History

Version History began during this release, so limited history is presented.

Component	Version	Date	Changed By	Description of Change
FCD	1.0	06/18/04	IPT	Initial Release
Schema	2.3	06/18/04	IPT	First Production Release
FCD	2.0	03/19/07	IPT	Include data publishing services; simplify data synchronization process.
Schema	3.0	03/19/07	IPT	Various improvements; disintegrated ("9 file") schema retirement; SSC adoption.
Schema	3.0	02/24/09	Windsor Solutions, Inc.	Added elements from the GeoRSS schema to allow for expression of a facility's location as a point, line, box (envelope), or polygon.
FCD	3.0	05/15/09	Windsor Solutions, Inc.	Updated the FCD for the Node 2.0 specification.
FCD	3.0a	5/28/2013	Windsor Solutions, Inc.	Clarify use of pipe symbol as multi-value parameter delimiter for only Node v1.1 Query/Solicit requests.

3.2 Flow Component Versions Currently Supported

The current component versions are listed below however, as per Exchange Network policy, the current and penultimate versions will be supported to ensure continuity of service during the transition to the new version.

Component	Version(s) Supported
FCD	3.0
Schema	3.0
DET	3.0
Schema User Guide	3.0

3.3   Business Rule Change History

Business Rule Change	Date of Change	Explanation (optional)
N/A		

4   Flow Summary Information

4.1   Flow Identification

Flow Name: Facility Identification

Flow Description:
Many environmental agencies have realized the importance of considering environmental issues using a holistic perspective of the impacts of regulated entities. To help support this form of environmental management these agencies have developed integrated information systems that allow them to access data about these facilities or sites along with contextual information regarding the potential affects they have on the environment; i.e., their environmental interests.

This diverse type of data is collected based on a wide variety of environmental regulations, and is often managed by a variety of local, state and federal agencies. These agencies can further benefit by sharing this key information so that each can incorporate a wider array and more comprehensive shared set of information within their analyses.

Flow Steward:
IPT Co-Chairs:
* David Smith, US EPA
* Dennis Burling, Nebraska DEQ

Flow Steward Contact Information:
Smith.DavidG@epamail.epa.gov; dennis.burling@nebraska.gov

4.2   Data Flow Overview

Partners can share this data either by synchronizing their databases with one another (e.g., to keep EPA's FRS system up to date with the facility data that is collected and reconciled by the many state and tribal agencies), or by publishing their data so that it can be used for more ad hoc exchange and analysis.

4.2.1   Data Synchronization

Data synchronization of Facility Identification data should occur by the incremental exchange of facility data on a periodic basis (e.g., weekly updates). This should be achieved by a partner providing XML payloads that contains <u>all</u> facility data for any facilities that have had <u>any</u> data modification since a requested 'change date'. For example, if one NAICS code is added for a facility, or the coordinate of a facility is changed, then the entire set of data for that facility must be exchanged. A single data service (GetFacilityByChangeDate) supports the exchange of the data in this manner.

This single approach will ensure all data modifications can be replicated with one exception. If a facility is deleted, then this event must also be communicated as part of this incremental

synchronization approach[1]. To support this, one additional data service is provided (GetDeletedFacilityByChangeDate) that allows a Partner to exchange a simple list of the key identification fields of each facility that has been deleted within the time period since the last incremental data exchange.

4.2.2 Dynamic Data Publishing

Data publishing is the process of making data available for general inquiry through a set of exposed web services using the Query and/or Solicit primitives. These data services allow a data requestor to employ a variety of filtering criteria (parameters) in any combination. By providing a variety of criteria, it is anticipated that much smaller payloads can be exchanged as more precise responses to requests are now possible.

To further allow a requestor to control the volume of data exchanged, the publishing data services provide four levels of granularity that can be provided within the response payload:
1. A count of the number of facilities that match the request
2. A 'list' of the facilities; (i.e., a few basic fields for each facility)
3. A slightly more expanded set of data, specifically including the environmental interests[2] for each facility
4. The full set of facility identification data including contacts and geographic information

By providing both a variety of data service parameters and a variety of payload data content these services will provide superior support for the anticipated increasing demand for more targeted and frequent exchange of this data across a broader set of Partners.

4.3   Flow Access and Security

NAAS has typically been the security mechanism for this exchange. However, NAAS security is not a prerequisite for this exchange. Partners must determine the security they will require, if any, for access to their data.

4.4   Flow-level Business Rules

Current Business Rules:
None (other than those documented in the Schema and DET).

Fault Follow-up Actions:
None

---

[1] Although most State systems and FRS have similar data models (as documented in the Facility Information Template for States, FITS I, and FITS II), there are a wide variety of ways such systems handle "deletes" of information, and even what "delete" means. As any veteran of facility data reconciliation can attest, many "deletes" are in fact "merges" of data from two duplicate facility records. In addition to merges, closures or other changes of regulatory status may, in some cases, produce "deletes".

[2] Environmental Interests are the environmental programs with which a given facility is involved (e.g. air emissions, surface water, ground water, and hazardous waste).

5 Schema Information

The Facility flow is supported by four schemas. The most complex services will return the simplest data set.

1. **Facility Details Schema:** This is the comprehensive schema. This schema contains all elements of the facility data standard, and some additions that have been part of the exchange historically, or which the IPT has adopted (e.g. Facility Status).

2. **Facility Index Schema:** This consists of only the highest level information about a facility. The primary elements included are the combination of elements that serve to uniquely identify the facility record (Originator, system, and ID), a URL to access a web site specific to the facility (e.g. Envirofacts or a State profiler application) and the physical location (address, coordinate).

3. **Facility Interest Schema:** This consists of the elements of the "Facility List Schema" plus each facility's set of environmental interests, containing:

   o The combination of elements that uniquely identifies the interest (partner name, system, ID).

   o A URL unique to that interest (if one exists) providing detail related to the interest in question.

4. **Facility Count Schema:** A simple schema whose sole purpose is to return the results of the GetFacilityCount_v3.0 service. It contains only one element, FacilityCountMeasure.

Please refer to the Schema User Guide for more details.

6   Data Service Information

6.1   Data Service Summary

The following table lists the seven data services available in the Facility Identification v3.0 exchange.

Service Name or Description	Service Types	Notes	Return Schema
GetFacilityCount_v3.0	Query	Partners may choose to implement this internally for costing purposes, so that over-large queries can be intercepted before they are run.	FACID_FacilityCount_v3.0.xsd
GetFacilityList_v3.0	Query or Solicit	Full set of parameters but returning on the lightest payload (FacilityIndex) to aid in performance.	FACID_FacilityIndex_v3.0.xsd
GetFacility_v3.0	Solicit	Full set of parameters, also including Change Date, and returning a payload based on the full schema.	FACID_FacilityDetails_v3.0.xsd
GetFacilityInterest_v3.0	Query	Full set of parameters but returning a payload of summary facility and environmental interest data (FacilityInterest) rather than the full schema, to aid in performance.	FACID_FacilityInterest_v3.0.xsd
GetFacilityByID_v3.0	Query	Get detailed facility data for one facility.	FACID_FacilityDetails_v3.0.xsd
GetFacilityByChangeDate_v3.0	Solicit	Used to support the creation and maintenance of a replica set of facility data across Partners (i.e., data synchronization).	FACID_FacilityDetails_v3.0.xsd
GetDeletedFacilityByChangeDate_v3.0	Solicit	This uses the FacilityIndex schema to return of basic identification data about each facility that has been deleted. This would only be used if a Partner is maintaining a replica set of facility data.	FACID_FacilityIndex_v3.0.xsd

6.2   GetFacilityCount_v3.0

Data Flow Name: FacID_v3.0

Data Service Type: Query

Data Service Parameters, Order, and Format:

Position	Name	Data Type	Occurrences	Max Length	Wildcards	Notes and Examples
1	Standard Environmental Interest Type	Text	0..∞	60		Return facilities having one or more of any of the standard types[3] included in the parameter array. Node v1.1 endpoints must use a pipe delimiter (i.e., '|') between values when multiple query values are present.
2	ZIP Code	Text	0. .∞	14	Starts with	The combination of the 5-digit Zone Improvement Plan (ZIP) code and the four digit extension code (if available) that represents the geographic segment that is a subunit of the ZIP Code, assigned by the U.S. Postal Service to a geographic location; or the postal zone specific to the country, other than the U.S., where the facility site is located. Will return all facilities in the provided zip code. If Zip+4 is provided, but the receiving node supports only 5 digit codes, treat the input parameter as a 5-digit code. Do not return facilities where zip code is null or unknown. Node v1.1 endpoints must use a pipe delimiter (i.e., '|') between values when multiple query values are present.
3	Tribal Land Code	Text	0..1	1		Code indicating whether or not the facility site is located on tribal land. *Allowable Values*: Y, N Return only facilities physically on tribal land. If the parameter is unsupported, return nothing.

---

[3] A list of standard environmental interest types can be found at the EPA [Environmental Data Registry](#).

Position	Name	Data Type	Occurrences	Max Length	Wildcards	Notes and Examples
4	Federal Facility	Text	0..1	1		Code indicating whether or not the facility is the property of the Federal Government. *Allowable Values*: Y, N If the facility or any interest has an ownership or operator type of "Federal", include the facility in the return set. If the parameter is unsupported by the receiving system, return nothing.
5	Facility Name	Text	0..1	80	Any use of '%' or '_'	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). Return all facilities where the primary name or any alias matches the supplied parameter. (Note: The returned data set may not include the matching name, since alternate names are not part of the list schema. The requesting party should address any possible confusion that results.)
6	Facility status	Text	0..1			Indicator of Active or Inactive, or other status that may be adopted as a part of the overall schema development. Return all facilities matching this parameter if supplied, or all if it is unsupported.
7	SIC Code	Text	0..∞		Starts with	The industry types that describe the business operations at the Facility. When searching using a SIC code, the system will find any facilities that are in that industry, even if the data about that facility is limited to the equivalent NAICS code(s). Return all facilities matching this parameter if supplied, or all if it is unsupported. A facility is a match if the facility or any interest lists a matching code--whether primary or not. The "Starts with" construct allows grouping of code sets, such as "primary metals"--this construct is supported by the underlying code set. Node v1.1 endpoints must use a pipe delimiter (i.e., '|') between values when multiple query values are present.

Position	Name	Data Type	Occurrences	Max Length	Wildcards	Notes and Examples
8	NAICS Code	Text	0..∞		Starts with	When searching using a NAICS code, the system fill find any facilities that are in that industry, even if the data about that facility is limited to the equivalent SIC code(s). Return all facilities matching this parameter if supplied, or all if it is unsupported. A facility is a match if the facility or any interest lists a matching code--whether primary or not. The "Starts with" construct allows grouping of code sets, such as "primary metals"--this construct is supported by the underlying code set. Node v1.1 endpoints must use a pipe delimiter (i.e., '|') between values when multiple query values are present.
9	City Name	Text	0..∞	60		The name of the city, town, village or other locality, when identifiable, within whose boundaries (the majority of) the facility site is located. This is not always the same as the city used for USPS mail delivery. Return only facilities where the city name of the physical location is an exact match to the supplied value. Some partners may record "nearest city" in this field, for facilities outside city boundaries. Actual placement with regard to boundaries may be unknown. Include these facilities if that is the case. Node v1.1 endpoints must use a pipe delimiter (i.e., '|') between values when multiple query values are present.
10	State	Text	0..∞	2		The U.S. Postal Service abbreviation that represents the state or state equivalent for the U.S. and Canada. Exact match is required. Node v1.1 endpoints must use a pipe delimiter (i.e., '|') between values when multiple query values are present.
11	County Name	Text	0..∞	35		The name of the U.S. county or county equivalent in which the facility site is physically located. Exact match is required. This value is more broadly supported than FIPS codes, but the FIPS table should be used for spellings--consistent with the data standard. Node v1.1 endpoints must use a pipe delimiter (i.e., '|') between values when multiple query values are present.

Position	Name	Data Type	Occurrences	Max Length	Wildcards	Notes and Examples
12	N Bounding Latitude	Number	0..1[4]			Northernmost coordinate of a bounding rectangle. Must be a valid decimal Latitude value. These 4 parameters define a box within which the returned facilities must reside based on a location coordinate for each facility. The operator is <= or >= as appropriate, with logical AND of all four conditions
13	S Bounding Latitude	Number	0..1			Southernmost coordinate of a bounding rectangle. Must be a valid decimal latitude value.
14	E Bounding Longitude	Number	0..1			Easternmost coordinate of a bounding rectangle. Must be a valid decimal longitude value.
15	W Bounding Longitude	Number	0..1			Westernmost coordinate of a bounding rectangle. Must be a valid decimal Longitude value.

Return Method: N/A

Payload Format: Facility Count Schema (FACID_FacilityCount_v3.0.xsd), contains a single <FacilityCountMeasure/> element.

Data Service-level Business Rules: N/A

Error Conditions and Fault Follow-up Actions: N/A

EN Header Usage: None.

Example Uses: Determine the number of facilities that match criteria – to determine if that exceeds the volume of data that is desired.

---

[4] If any box parameter is specified (positions 12-15), then all four box parameters are required.

6.3   GetFacilityList_v3.0

Data Flow Name: FacID_v3.0

Data Service Type: Query or Solicit

Data Service Parameters, Order, and Format:
The parameters as identified for the GetFacilityCount service above. Please refer to that listing.

Return Method: N/A

Payload Format: Facility Index Schema (FACID_FacilityIndex_v3.0.xsd)

Data Service-level Business Rules: N/A

Error Conditions and Fault Follow-up Actions: N/A

EN Header Usage: None.

Example Uses: Populate a catalog, allow users to browse and "drill down" as needed; explore available data at a summary level.

6.4   GetFacility_v3.0

Data Flow Name: FacID_v3.0

Data Service Type: Solicit

Data Service Parameters, Order, and Format:

The parameters as identified for the GetFacilityCount service above, but including the following additional parameter.

Position	Name	Data Type	Occurrences	Max Length	Wildcards	Notes and Examples
16	Change Date	Date	0..1			Date since when any data element of a facility has been modified. Response will include all facilities that have changed on or after this date.

Return Method: N/A

Payload Format: Facility Detail Schema (FACID_FacilityDetails_v3.0.xsd)

Data Service-level Business Rules: N/A

Error Conditions and Fault Follow-up Actions: N/A

EN Header Usage: None.

Example Uses: Restrict data set to avoid oversize files during synchronization. Well targeted and refined queries (e.g. agreed regular exchanges).

6.5   GetFacilityInterest_v3.0

Data Flow Name: FacID_v3.0

Data Service Type: Query or Solicit

Data Service Parameters, Order, and Format:

The parameters as identified for the GetFacilityCount service above. Please refer to that listing.

Return Method: N/A

Payload Format: Facility Interest Schema (FACID_FacilityInterest_v3.0.xsd)

Data Service-level Business Rules: N/A

Error Conditions and Fault Follow-up Actions: N/A

EN Header Usage: None.

Example Uses: Get further details regarding a facility that was returned within a Facility Index XML document.

6.6   GetFacilityByID_v3.0

Data Flow Name: FacID_v3.0

Data Service Type: Query

Data Service Parameters, Order, and Format:

Position	Name	Data Type	Occurrences	Max Length	Wildcards	Notes and Examples
1	Facility Site Identifier	Text	1..1	N/A	N/A	The unique identifier allocated to the facility by the originating partner within the source database system.
2	Originating Partner Name	Text	1..1	N/A	N/A	The name of the partner that originally provided the exchanged facility site or environmental interest data.
3	Information System Acronym Name	Text	1..1	N/A	N/A	The abbreviated name that represents the name of an information management system for an environmental program.

Return Method: N/A

Payload Format: Facility Detail Schema (FACID_FacilityDetails_v3.0.xsd)

Data Service-level Business Rules: N/A

Error Conditions and Fault Follow-up Actions: N/A

EN Header Usage: None.

Example Uses: Get full details regarding a facility that was returned within a FacilityIndex or FacilityInterest XML document.

6.7   GetFacilityByChangeDate_v3.0

Data Flow Name: FacID_v3.0

Data Service Type: Solicit

Data Service Parameters, Order, and Format:

Position	Name	Data Type	Occurrences	Max Length	Wildcards	Notes and Examples
1	Change Date	Date	1..1	N/A	N/A	Date since when any data element of a facility has been modified. Response will include all facilities that have changed on or after this date.
2	Originating Partner Name	Text	0..1	N/A	N/A	The name of the partner that originally provided the exchanged facility site or environmental interest data.
3	Information System Acronym Name	Text	0..1	N/A	N/A	The abbreviated name that represents the name of an information management system for an environmental program.

Return Method: N/A

Payload Format: Facility Detail Schema (FACID_FacilityDetails_v3.0.xsd)

Data Service-level Business Rules: N/A

Error Conditions and Fault Follow-up Actions: N/A

EN Header Usage: None.

Example Uses: Get all facility data so that a replica of that data can be retained and synchronized.

6.8   GetDeletedFacilityByChangeDate_v3.0

Data Flow Name: FacID_v3.0

Data Service Type: Solicit

Data Service Parameters, Order, and Format:

Position	Name	Data Type	Occurrences	Max Length	Wildcards	Notes and Examples
1	Change Date	Date	1..1	N/A	N/A	Date since when any data element of a facility has been modified. Response will include all facilities that have changed on or after this date.
2	Originating Partner Name	Text	0..1	N/A	N/A	The name of the partner that originally provided the exchanged facility site or environmental interest data.
3	Information System Acronym Name	Text	0..1	N/A	N/A	The abbreviated name that represents the name of an information management system for an environmental program.

Return Method: N/A

Payload Format: Facility Index Schema (FACID_FacilityIndex_v3.0.xsd)

Data Service-level Business Rules: N/A

Error Conditions and Fault Follow-up Actions: N/A

EN Header Usage: None.

Example Uses: Get all deleted facilities so that a replica of that data can be retained and synchronized.