Environmental Information exchange Network

EN2015

Envirofacts API

Cory Wagner, *US EPA*

2015 Exchange Network National Meeting

Supporting the Business of Environmental Protection

September 29–October 1, 2015
Sheraton Philadelphia Society Hill Hotel
Philadelphia, Pennsylvania

http://www.exchangenetwork.net/en2015

ABSTRACT

Envirofacts is the EPAs One-stop shop for environmental data. This presentation will discuss the current implementation of the Envirofacts API along with the future plans to modernize it

What is Envirofacts?

- Warehouse of Integrated Environmental Data
- Suite of Web Tools
- Enables users to **find and map** environmental information
- Helps **analyze** data to meet business and information needs
- Provides data and information for **EPA** and non-EPA (**Public**) use

Envirofacts Facts

EF Database Stats

3488 Tables

(program office databases, user tables, and application-specific tables)

500 Gigs

Usage Stats

12 Million Hits Annually

against the EF Website

15 Million API Hits Annually

against the EF REST API Services supporting external data utilization

Users

12 GIS Applications

30+ EPA customers

(18 relying on EF for public data access)

10 EF data streams

with other (non-EPA Agencies)

4 Mobile Applications

supported by EF

3 EPA Registries

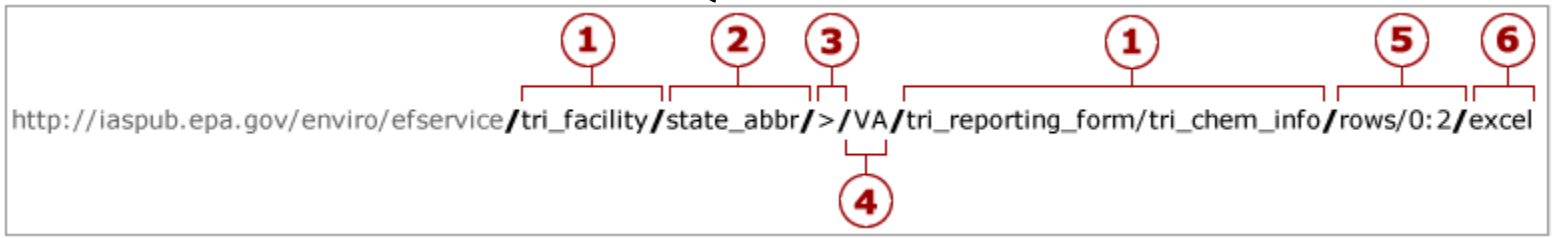
supported by EF

Envirofacts API

- Envirofacts has a RESTful data service API to all of its internal data holdings
  - This functionality provides a user of Envirofacts with the ability to query any table through the use of a URL
  - The default output is in XML, which can be utilized in other applications, as well as tools such as Microsoft Excel or Access. Output options of JavaScript Object Notation (JSON), CSV, and Excel can be requested in the URL
  - Services are documented in the Environmental Dataset Gateway (EDG)
  - Services will be made available through the E-Enterprise Portal via the Shared Services Catalog and through Developer Central

Example of the Envirofacts API

Point out that the current API is SQL based and that we



1. **Table Name** – At least one table name is required. Up to three table names can be entered.
2. **Column Name** – This is an optional entry. Enter a column name and value to limit the results.
3. **Operator** – This is an optional entry. This parameter allows users to pass in an operator with the query. Default output is "=" and does not require an operator, but users can enter "<", " >", "!=", "BEGINNING", "CONTAINING", operators as well via the URL.
4. **Column Value** – This is an optional entry. The column value is queried against the database without modification.
5. **Rows** – This is an optional entry. Specify the rows to display by entering 'rows/ <first_row> : <last_row>' Results numbering starts at 0. So to get the first five hundred rows enter rows/0:499
6. **Output Format** – This is an optional entry. The default output is in XML; however, output options of JSON, CSV or Excel can be requested in the URL.

EF API Strengths and Limitations

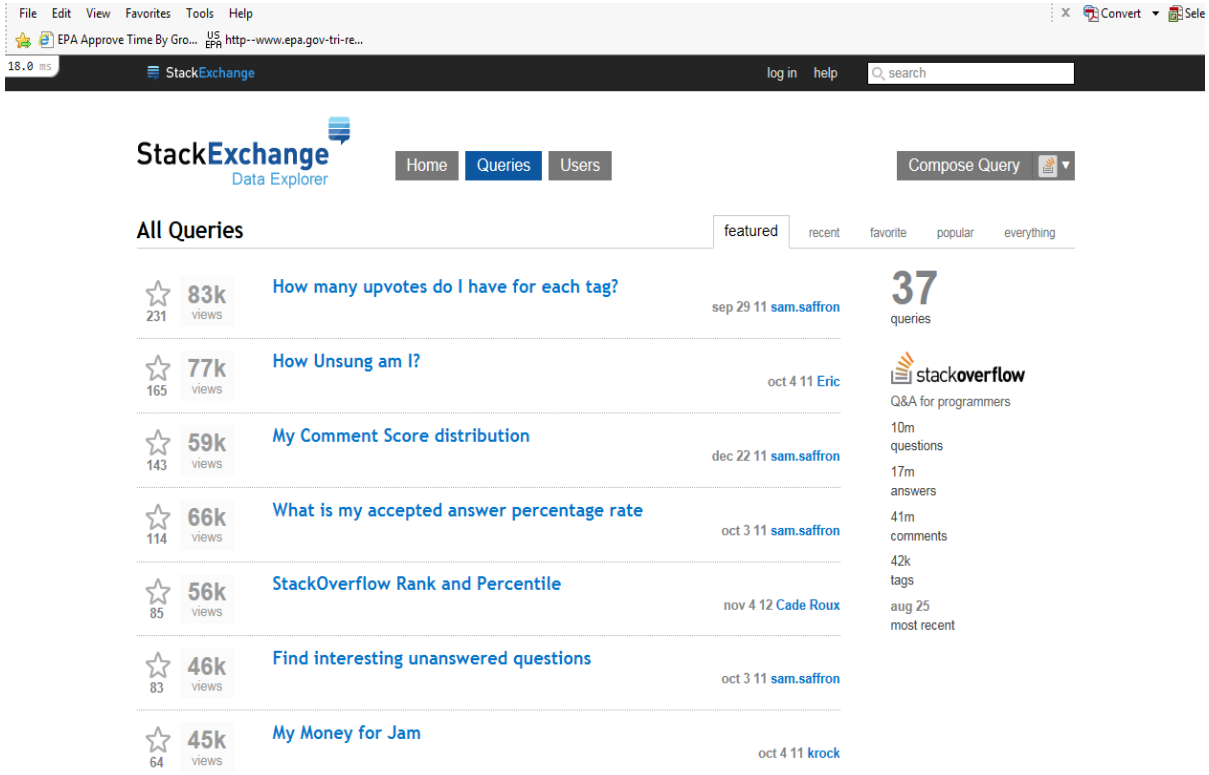
- Strengths:
  - Simple and well-documented
  - Access multiple data sources without the need of user database connections.
  - Output can be easily used with desktop applications such as Excel or Access, or used in creating Web mash-ups.
- Limitations:
  - Can only query up to three tables (out of hundreds)
  - Output is limited to 10000 rows of data at a time
  - Joins can be very complex

Proposed Improvements to the EF API

- The row and join restrictions on the EF API limit its utility
- As part of a modernization effort, OEI is looking at the services with an eye on how to improve them
- Proposal: Create an Advanced Query Builder interface for Envirofacts
  - The AQB:
  - Provides an easy web-based access to the latest and greatest Envirofacts data dumps.
  - Gives users the ability to bulk download data via CSV format and an API endpoint to access the query results.
  - Allows testing, running, editing, sharing, forking and permalinking to public queries against Envirofact's corpus of data with a simple, syntax-highlighted web user interface.
  - Can be used as a permanently linkable tool for teaching general SQL and relational database concepts.
  - Would be hosted in a cloud environment so that it can be speedy, scalable, and always available.

AQB Details

- The AQB is based on the Stack Exchange Data Explorer (http://data.stackexchange.com/)
- Open Source tool built on an ASP.NET MVC software Stack in a Windows Azure environment

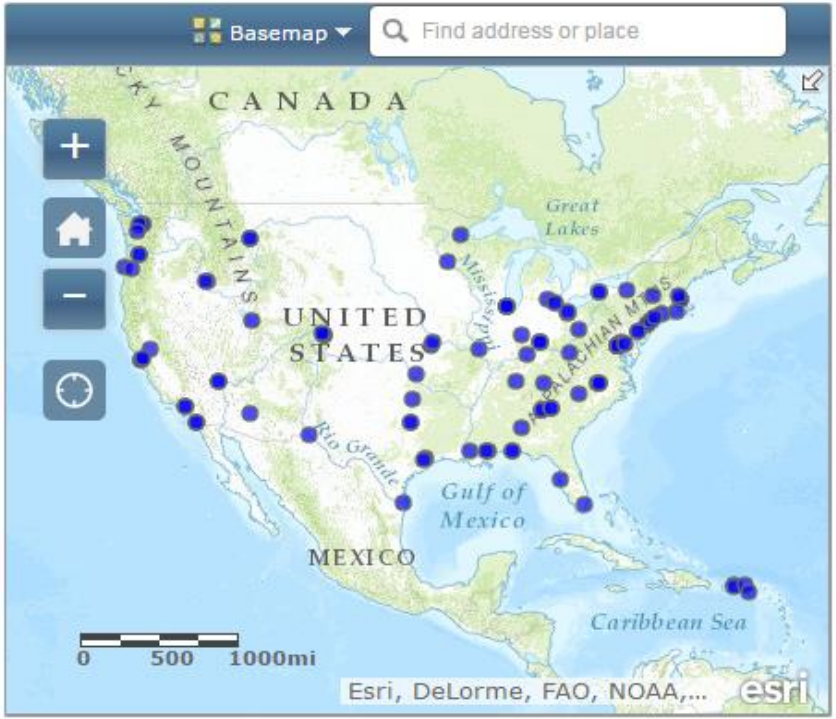


AQB Features

- Ability to feature select queries and show popular queries
- Ability for anonymous users to like/star a query
- Ability to choose between various systems (e.g. BR, RCRAINFO, CERCLIS, ECHO [+ Core FRS and Location Information tables] etc.)
- Ability to assign plain English titles, descriptions and tags to queries
- Robust search functionality allowing users to search for titles, descriptions, tables, columns and tags
- Comprehensive help and tutorial sections incorporated into the UI.
- Ability to fork queries so that users can re-purpose existing queries.
- Ability to view results and download results in CSV format
- Ability to reference results in JSON format through a REST API endpoint
- Auto detection lat/long fields and displays the data on a map when applicable
- Incorporates existing business rules to prevent incorrect table joins when possible.

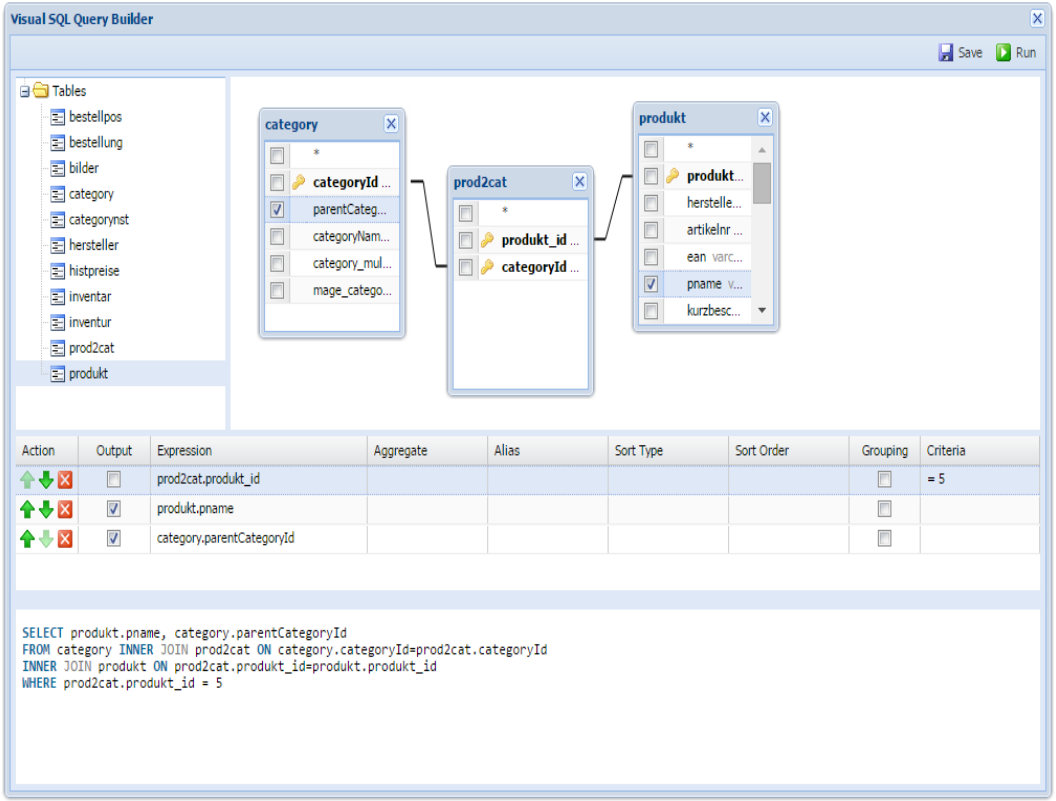
AQB Features

- Auto detection lat/long fields and displays the data on a map when applicable
- Utilize EPA's Geoplatform



AQB Features

- Visual SQL Query Builder to allow users to:
  - drag tables
  - create an alias
  - order/sort/aggregate
  - support grouping
  - allow for the addition of criteria
  - allow for quick creation of table joins.



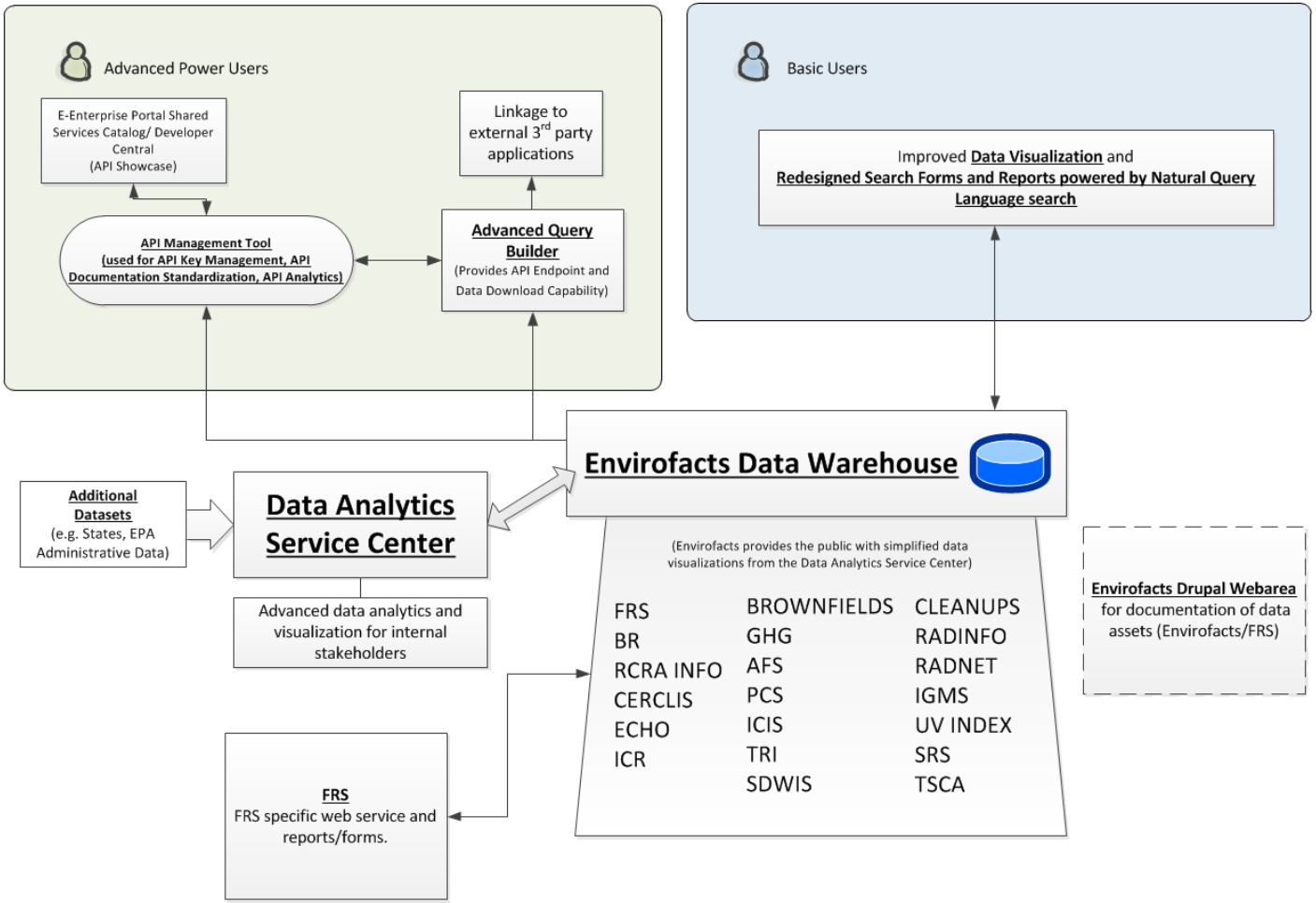
API Management Tool

- Also looking at API Management Tools
- Likely incorporated into first release of the AQB
- May require registration
- Provides standardized documentation and more robust analytics
- This will provide us with a better assessment of user needs

AQB Debut?

- Plan to have a version of the AQB available in 2016
- Along with other enhancements to Envirofacts
  - Unstructured data inclusion
  - Natural Language Querying allowing a refined Google bar interface for the basic interface
  - Refined look and feel to incorporate a more Red Fin like facility view as well as data visualizations to replace some of the tabular results
  - And maybe more!

Envirofacts Redesign Schematic



Envirofacts Contacts

General: Bill Muldrow
muldrow.bill@epa.gov
202.566.0692

AQB/API: Andrew Yuen
yuen.andrew@epa.gov
202.566.1308

Cory Wagner
wagner.cory@epa.gov
202.566.1555

www.epa.gov/enviro