Using Web Services to Publish Data

2010 ASTSWMO Hazardous Waste Managers Conference

Dwane Young, EPA/OSWER/ORCR

June, 2010

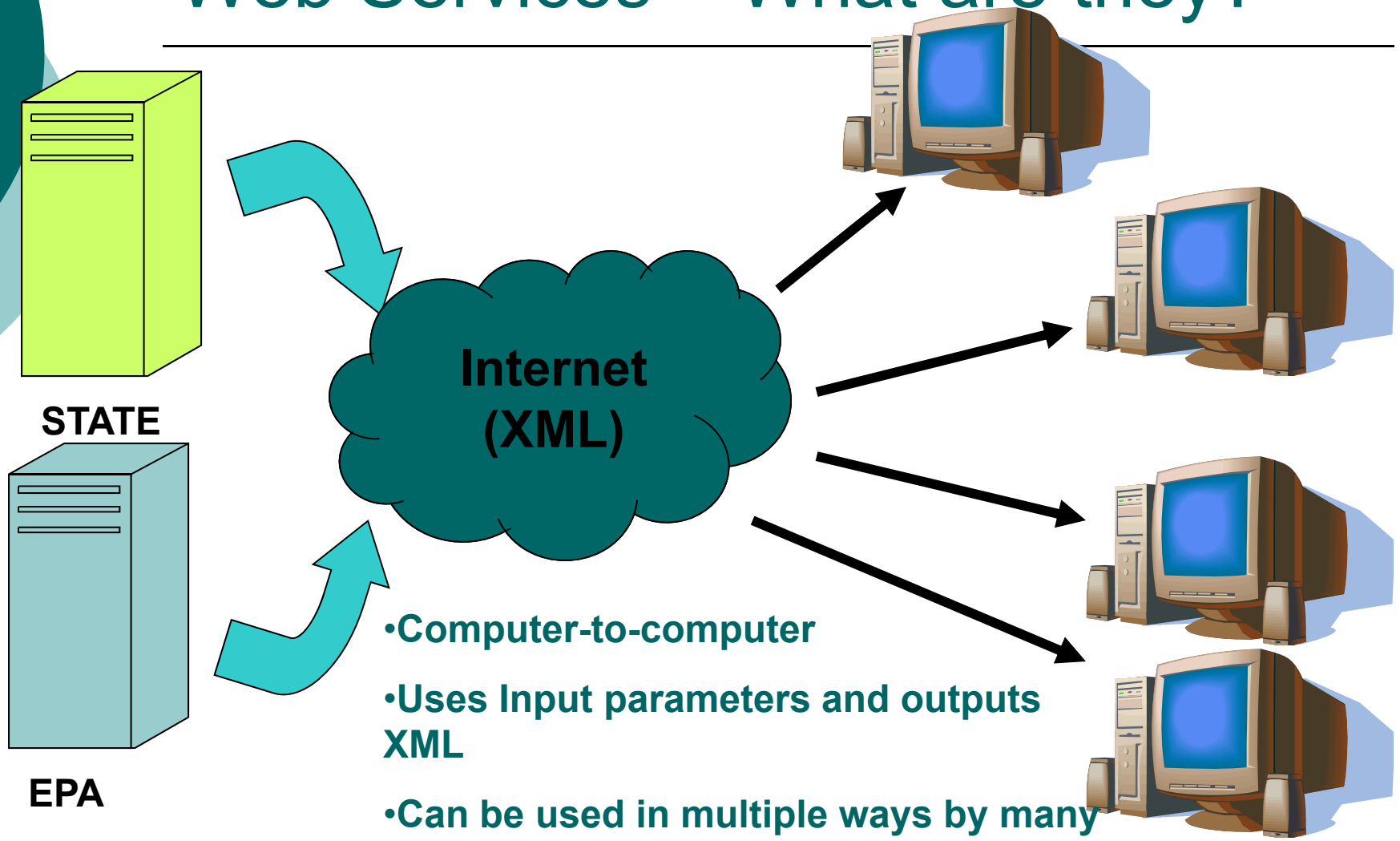
Overview

- EPA Program Offices have enhanced data access by implementing public-facing web services to allow for better access to programmatic data
- This approach allows for services to be re-used across many different applications, and also allows for the data to be integrated in ways that it hasn't been before

Web Services: Described

- Standardized way for applications to communicate
- Platform independent
- Programming language independent

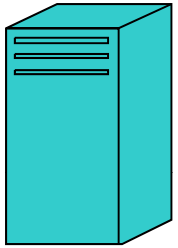
Web Services – What are they?



- Computer-to-computer
- Uses Input parameters and outputs XML
- Can be used in multiple ways by many applications

Web Services – An Example

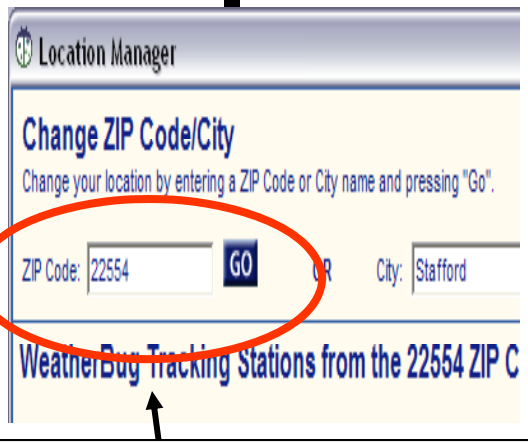
Weather Network

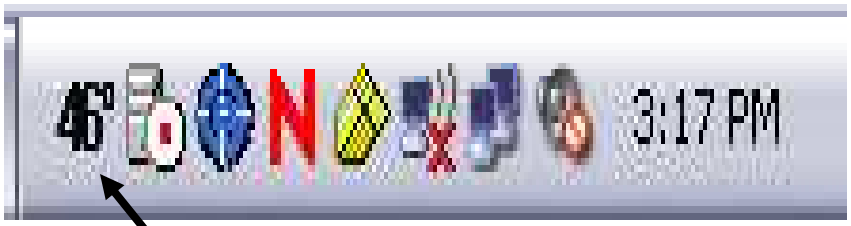


Weatherbug is an example that many are familiar with

This is all done via a Web Service

Returns XML





Input Parameter: Zip Code

Weatherbug translates XML into information for the task bar

How RCRAInfo Uses Web Services

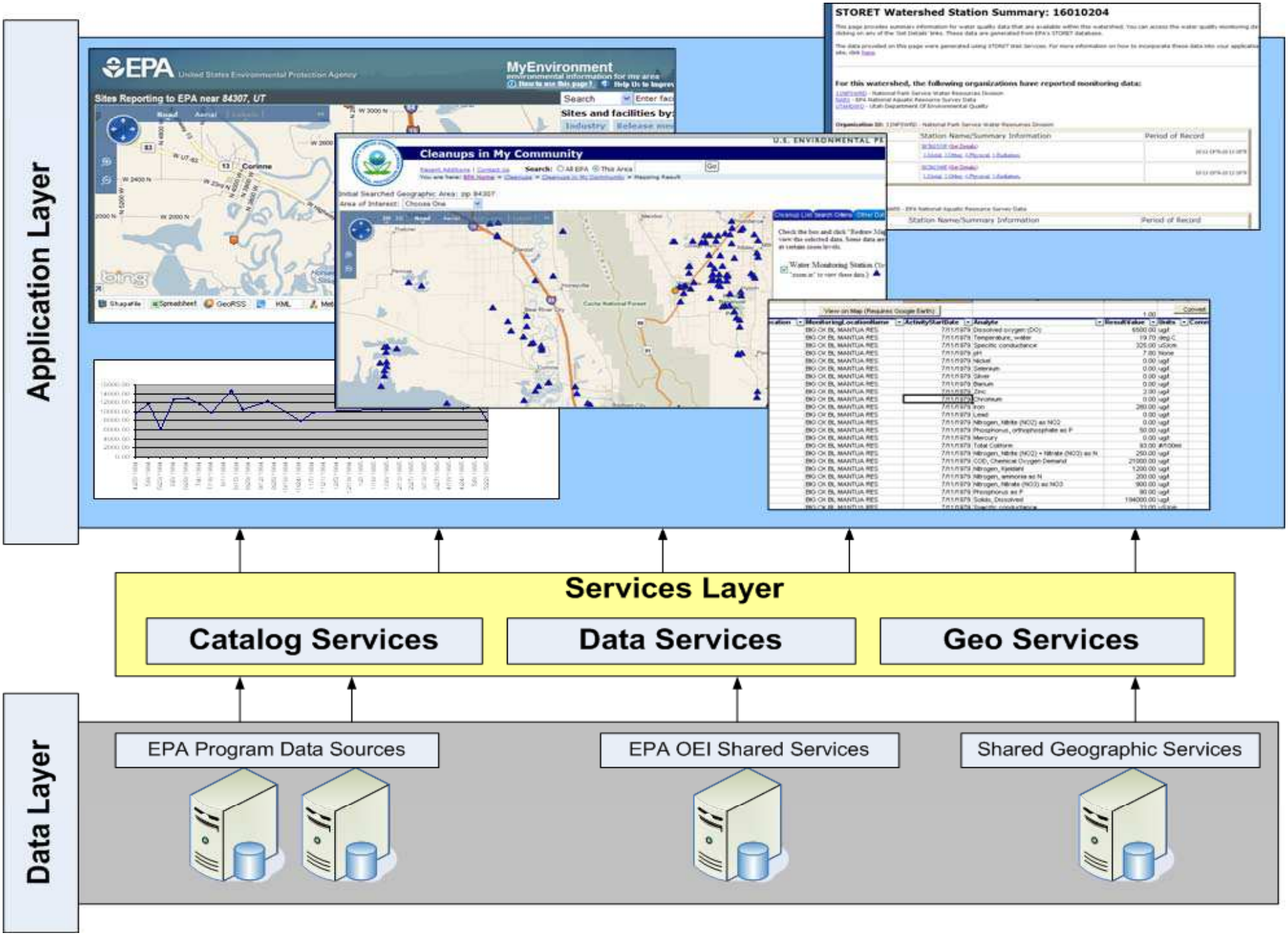
- Accept data submissions via Exchange Network
- Allow users to download load status reports
- Communicate with CDX Node
- Provide data synchronizing and retrieval functionality

History of data exchange at EPA

- Historically, web services have been used at EPA to publish:
  - Geospatial data
  - Exchange Network (EN) data flows
- EN utilizes XML schema for EPA-trusted partner data exchange of regulatory flows
- EPA also publishes data from Envirofacts and program-specific databases via queries, reports, text files, and downloads

The future: using web services

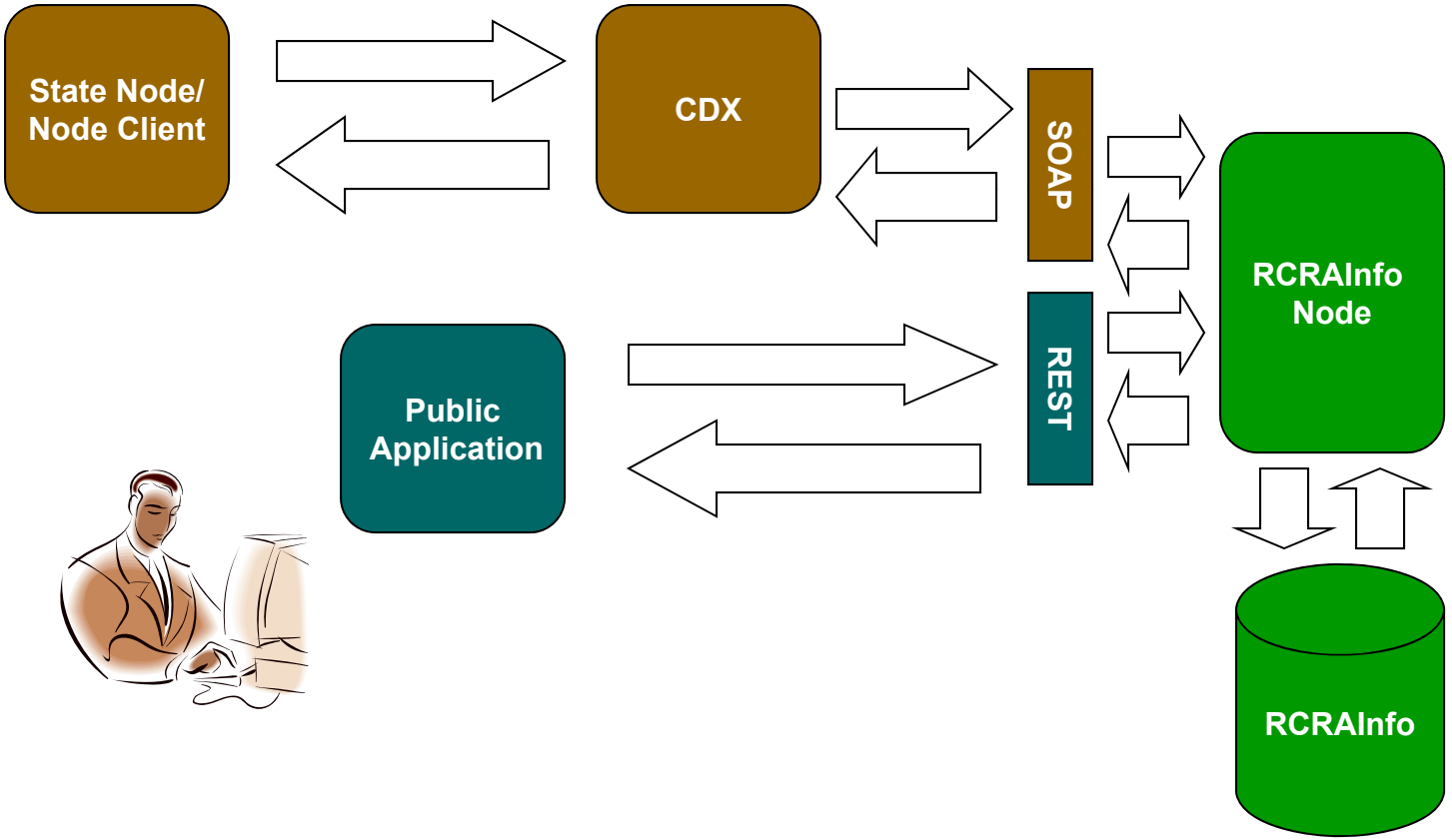
- Goal: reusable, interoperable, standards-based tools, services, and components implemented '*a la carte*' to promote discovery/access
- Web services promote public-facing data access, as well as trusted-partner access
- Support re-use of data in a variety of types and services, including OGC and spatial web services (WFS/WMS, KML, GeoRSS)


Application Layer
EPA United States Environmental Protection Agency
MyEnvironment
Sites Reporting to EPA near 84307, UT
Sites and facilities by:
Industry
Cleanups in My Community
STORET Watershed Station Summary: 16010204
For this watershed, the following organizations have reported monitoring data:
Station Name/Summary Information
Period of Record
Services Layer
Catalog Services
Data Services
Geo Services
Data Layer
EPA Program Data Sources
EPA OEI Shared Services
Shared Geographic Services

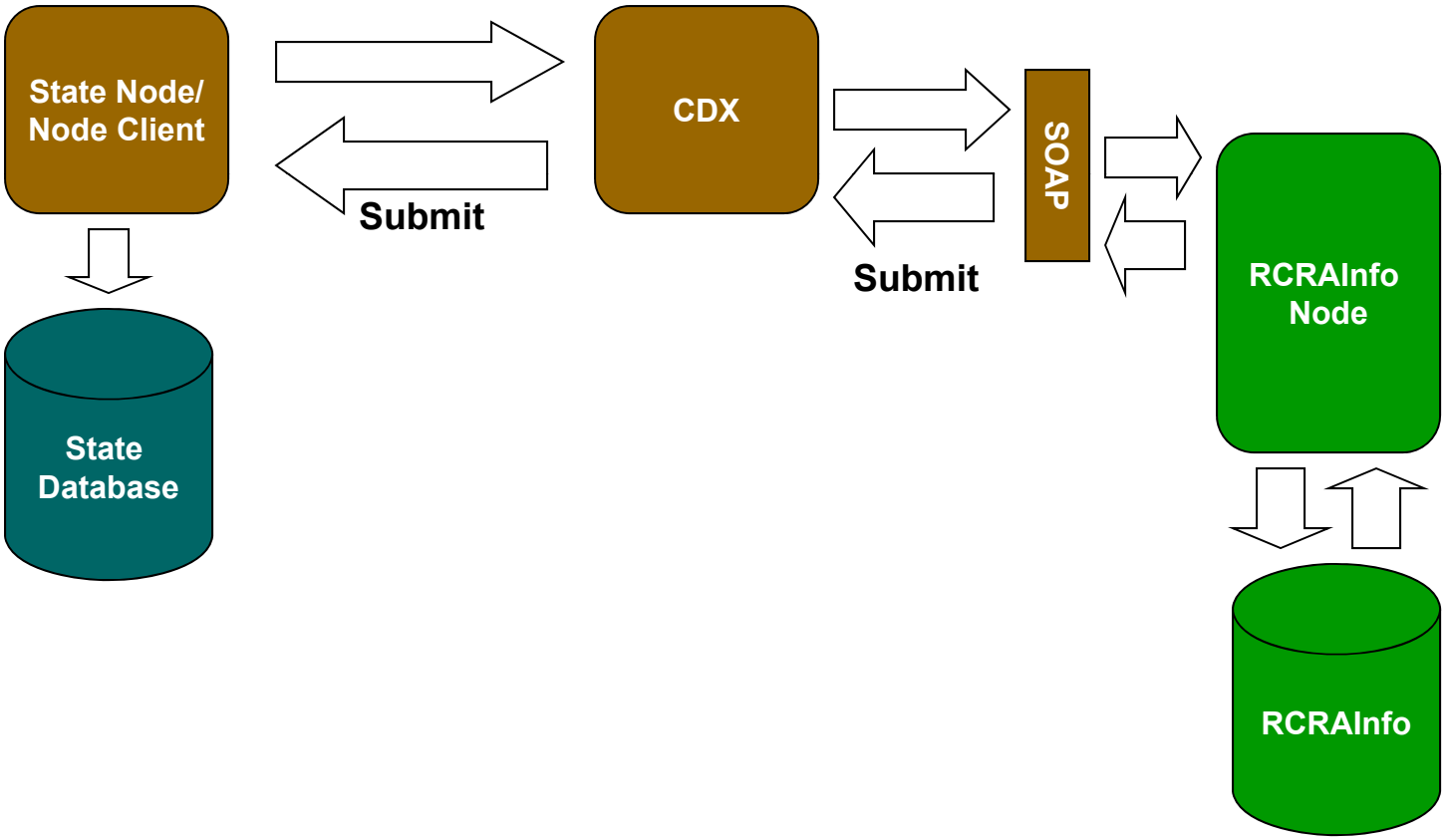
The OSWER Approach

- Exchange Network (EN) web services
  - Simple Object Access Protocol (SOAP)
  - Submit data to RCRAInfo
  - Query/Solicit data from RCRAInfo
- RCRAInfo public facing web services
  - Representational State Transfer (REST)
  - Retrieve public data from RCRAInfo

Query and Public REST Service



Solicit Services



Why we did it this way?

- Have both Public and Exchange Network data sharing needs:
  - The REST services provided a better public interface
  - The public data needed to be outside a username/password protected area
  - Exchange Network use case for States to sync their databases with data contained in RCRAInfo

Services that EPA is considering

- GetHDDataByHandlerID (both Public and EN)
  - Retrieve Handler data from the Handler module by providing a Handler ID (a similar service will exist for all modules)
- Syncing Web Services (EN Only)
  - Used to sync state database with EPA databases
- Functional Services
  - GetGNMaxSequenceNumber (Public and EN)
  - GetGNNewHandlerID (EN Only)
- Catalog (Data Summary) services (Public and EN)

Web Services Schedule (Published)*

May 2010	Begin web services evaluation
Oct 2010	Initial web services implementation
Nov 2010	Begin testing with partners
March 2011	Finalize testing
Sept 2011	Roll-out final services

*We will likely roll-out some services earlier than others (for example the current GetHandlerByID REST service will likely go to production before the end of this September 2010.

Demo

- Using existing web services, we'll show how a simple spreadsheet can be used to pull RCRA data and water monitoring data together, without having to know anything about the underlying data structure