Design of a cloud-based system to integrate and share data for environmental governance and sustainability

Yu-chi Chu and Su-mei Huang
Dept. Environmental Monitoring & Information Management, EPA, TAIWAN

行政院環境保護署
Environmental Protection Administration
Executive Yuan, Taiwan, R.O.C.

Introduction

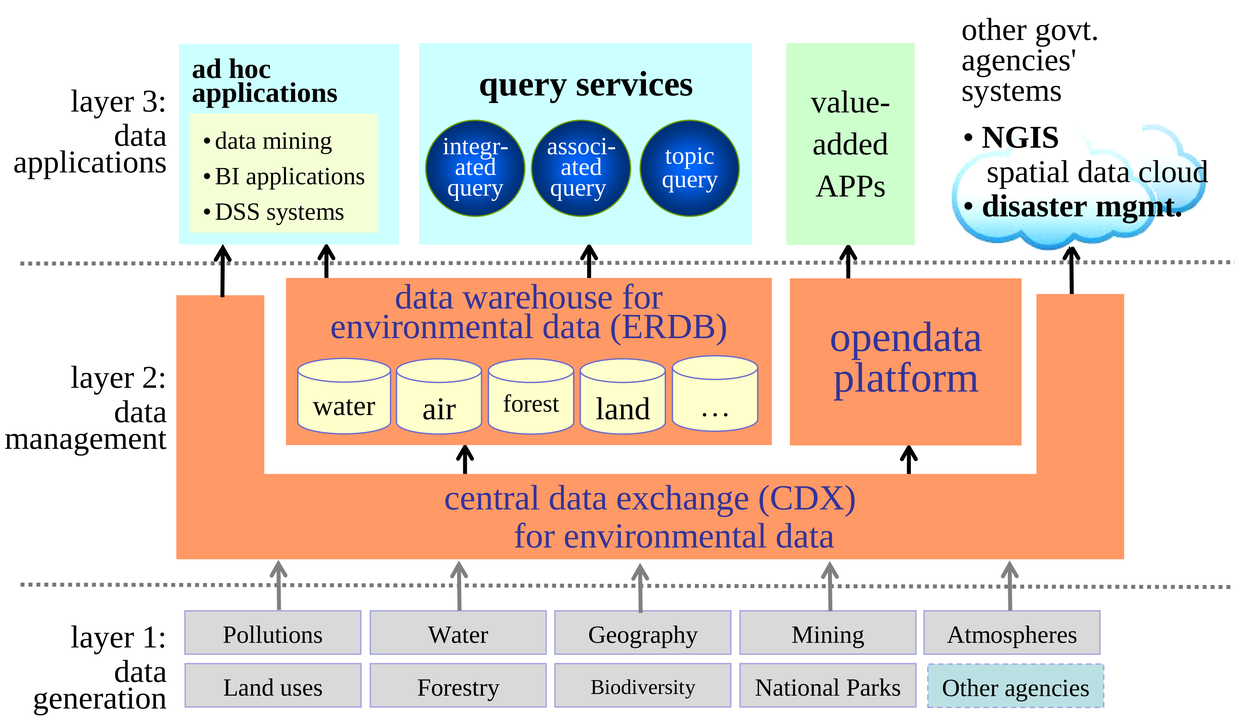
For making sensible, justifiable, and legally correct decisions, both government agencies and private sectors need detailed information regarding the current state of the environment and ongoing developments. Currently, it is very difficult to share environmental data since the information typically resides on geographically disparate and heterogeneous databases (systems).

These systems often do not facilitate access by secondary users and frustrate attempts to draw data together to form a more comprehensive understanding of environmental conditions and actions. Therefore, there is a major demand for appropriate systems and adequate tools to provide integrated information for managing the issues of environmental governance and sustainable development.

Missions

EnviroCloud project brought together 10 partners from 5 different federal agencies of Taiwan central government. Its aim is the requirement analysis and top level design of a comprehensive and cloud-based system which will collect, integrate, and share environmental information. The system has three missions:

1. establish a national wide data collection and gathering process and mechanism,
2. integrate and analyse collected data, transform the data to meaningful information, moving from information to insight,
3. develop ad hoc applications and disseminate environmental data to empower and support the decision making for environmental governance across agencies.



. *Figure 1: The architecture of* ***EnviroCloud***

System Architecture

The top level design of **EnviroCloud** project is three-layer architecture. Figure 1 depicts the architecture of the proposed method. We briefly explain the functions and its major components of each layer as follows.

1. **Data generation and collection**. In this layer, we select the sources of data and information that we consider relevant to the environment and natural resources, for example, atmosphere, air quality, water resource and quality, geology, forest and biodiversity, etc. These data and information are generated by participating partners that play as the "data publishers". They are also in charge of data quality assurance and provide metadata for the data they published.

2. **Data management**. This layer consists of three components that collaborate with each other to fulfil the missions of data management such as data transformation, data exchange, and data store. Firstly, the central data exchange (CDX) enables participating partners to work with stakeholders - including local governments to perform streamlined, electronic submission of data (machine to machine) via the Internet. Secondly, the opendata (OD) platform accesses metadata associated with the data that are allow to be used by general public from CDX. Components and modules of OD are maintained at opendata.epa.gov.tw. Thirdly, we will build a data warehouse which is dedicated to integrate environmental and natural resources data (called ERDB). The data warehouse will provide a foundation and play a crucial role for the operations of data analysis to next layer.

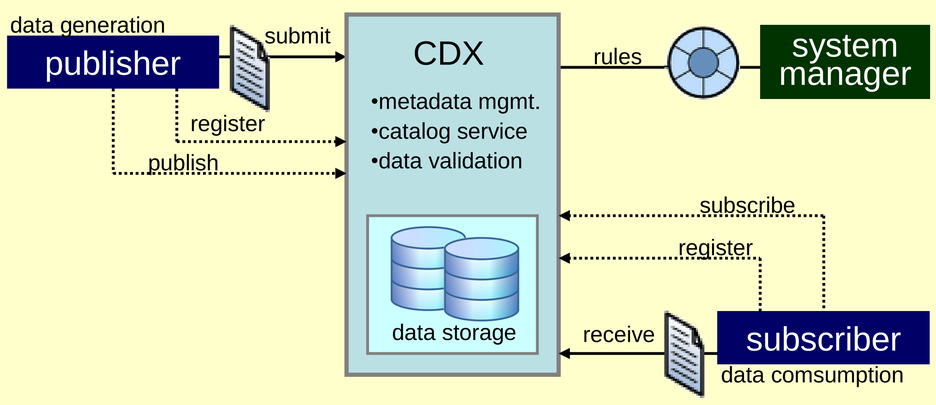
3. **Data applications and dissemination**. In this layer, we will develop a number of ad hoc and specific applications ranging from topic-oriented query systems to applications of business intelligence, as well as big-data related technologies. We encourage academic institutes and industries to "use" the data offered by opendata platform to develop software applications that might help environmental governance and sustainable development. We are also highly connected with the systems that are developed by some other government agencies, such as national geographic information system (NGIS) and disaster management systems. Therefore, EnviroCloud can benefit not only benefit participating partners but also other government agencies in the decision making process for environmental governance and sustainable development.

Publish-Subscribe Model for Data Exchange

Because the environmental data is collected by various government agencies, CDX provides a publish-subscribe solution for effective information delivery among these agencies. CDX enables Taiwan EPA and participating agencies to work with stakeholders - including local governments - to enable streamlined, electronic submission of data via the Internet. CDX also provides the capability for submitters to access their data through the use of web services.

Data that exchange through a publish-subscribe approach requires the senders to publish data without explicitly specifying recipients or having knowledge of intended recipients. Similarly, receivers must receive only those data that the subscriber has registered an interest in. Figure 2 illustrates publish and subscribe functionality of CDX. It provides participants with the ability to:

- Submit and receive data through one centralized point, in a variety of formats including Webs Forms, XML, or flat-file
- Utilize publishing services to share information collected by other stakeholders and exchange data with target systems using web services



. *Figure 2: The publish and subscribe functionality of CDX*

Opendata Platform

- Taiwan EPA views environmental data as a strategic asset to protect our environment. For bring tangible benefits to public, we are on track to fulfil the requirement of the opendata policy proposed by Executive Yuan, Taiwan central government in 2013.
- There are currently over 65 datasets on the opendata platform including hourly air quality monitoring data, UVI values, water monitoring data, etc. The datasets are in various formats such as JSON, XML, and KML for spatial data. We have taken a number of measures to encourage use of the datasets already published. For ensuring the data quality, we also establish a working group to oversee and provide input to the opendata efforts.

Future Works

EnviroCloud has been implementing a number of software systems and operation procedures. In the future, we will focus on the implementation of ERDB, a consolidated data warehouse system, to collect and materialize the data from CDX and other resources. In addition, we are currently investigating to employ big data technologies to analyse the vast amount data, both in structured and unstructured formats, collected from participant agencies and stakeholders.

Contact:

Phone: +886-2-23117722 ext. 2300, 2340
Email: ycchu@epa.gov.tw, sumhuang@epa.gov.tw
Web: http://www.epa.gov.tw
Addr: 81, Section 1, Zhonghua Rd, Taipei, 10042, Taiwan (R.O.C.)