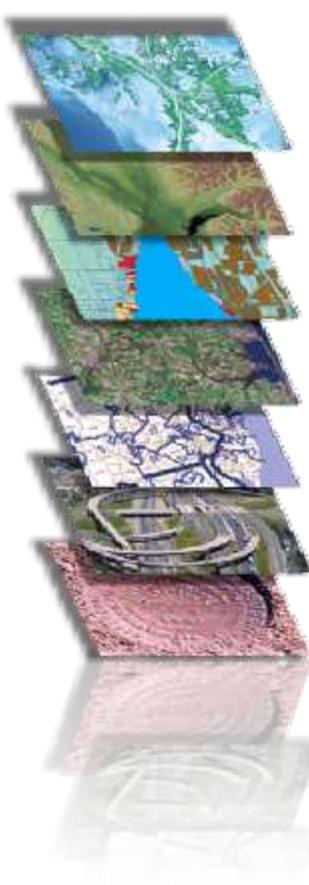




# Background on GeoCloud Sandbox Initiative

---

- Initiated as an Architecture and Technology Working Group activity in December 2009
- Five-year project to test and monitor externally-hosted Cloud data and service solutions for the geospatial domain, to support the Geospatial Platform activity
- Initiative Sponsors:



**FGDC.GOV**  
FEDERAL GEOGRAPHIC DATA COMMITTEE



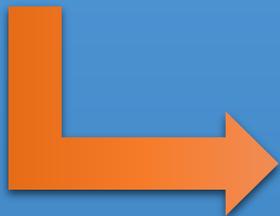
# Cloud Community Platform Initiative

## – *Mission Needs*



### Business Drivers

- Growing pool of identified agency applications seeking cloud benefits
- Quantify savings, identify risks, ease migration



### Platforms

- Construct a discrete subset of secure supporting platforms
- Develop processes for rapid application migration
- Leverage certifications and platform commonalities
- Develop chargeback mechanisms for self-sustaining funding



### Best Practices

- Capture lessons learned
- Document techniques, issues and solutions
- Document cost models and benchmarks
- Build a cloud platform support community
- Maintain platforms, scripts and updates for client agencies



# GeoCloud Programmatic Goals, Objectives, Activities and Outcomes



## Architectures

- Identify requirements-driven solution architectures and platforms for various sized deployments of geospatial data and services

Acquire, compose, document, and deploy reference platform cloud that support Geospatial Platform standards

## Cost Models

- Document and assess cost models to support scalability, reliability, and redundancy

Monitor costs, loads, issues and options in support of OMB IT project document guidance

## Certification

- Expedite FISMA (security) certification and accreditation for agency adoption of packaged solution architectures

Certify Geospatial Solution Packages to facilitate re-use

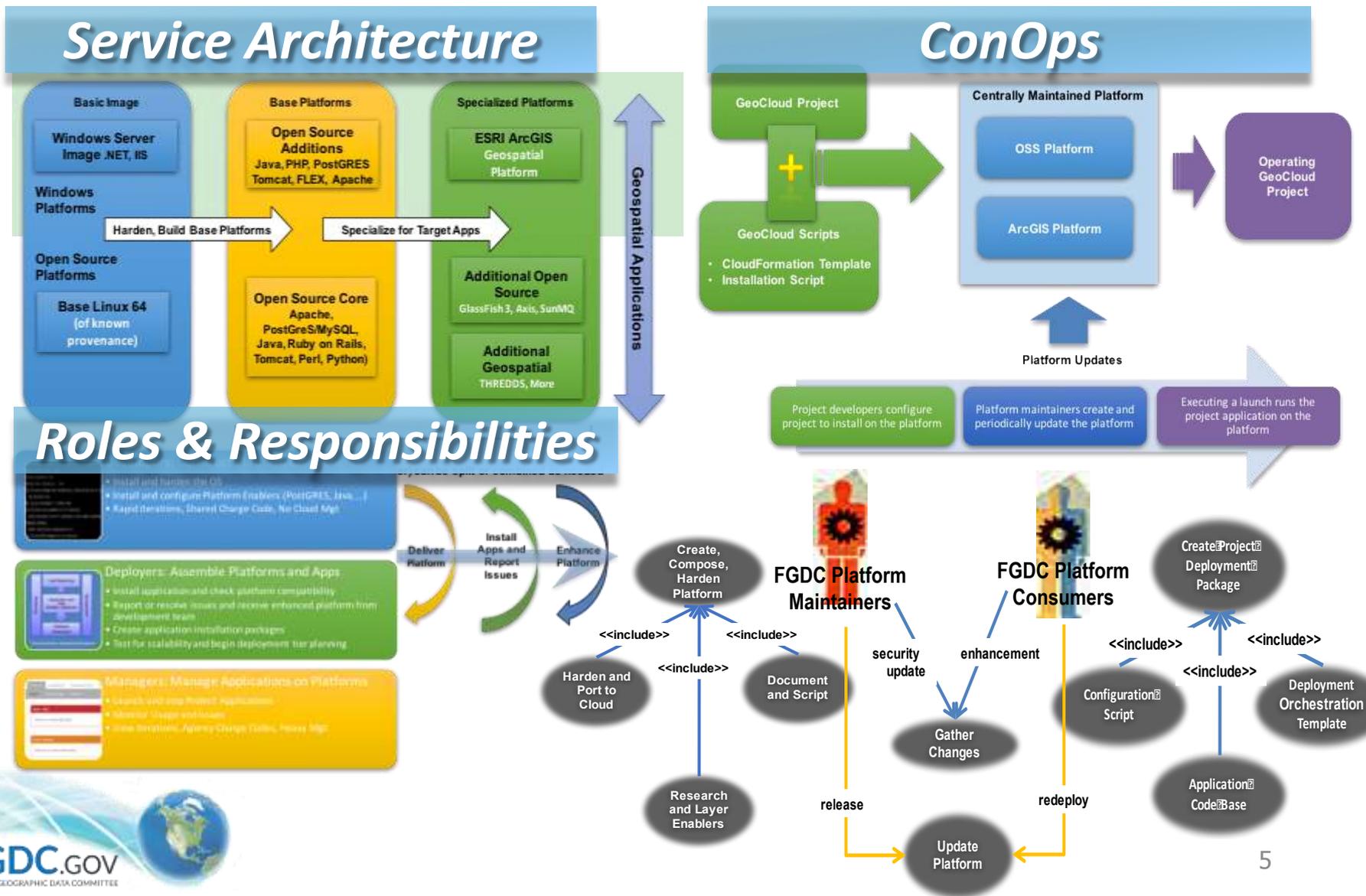
## Comparisons

- Support and collect cost comparison information from agencies for existing and externally-hosted Cloud solutions
- Document lessons learned and best practices

Document Best Practices and guides to agencies on adoption of geospatial Cloud infrastructure

# GeoCloud Outcomes

## – Development of a Cloud Services Framework



# GeoCloud Technical Goals and Objectives

## – *Community Platform Sandbox*

### Goals

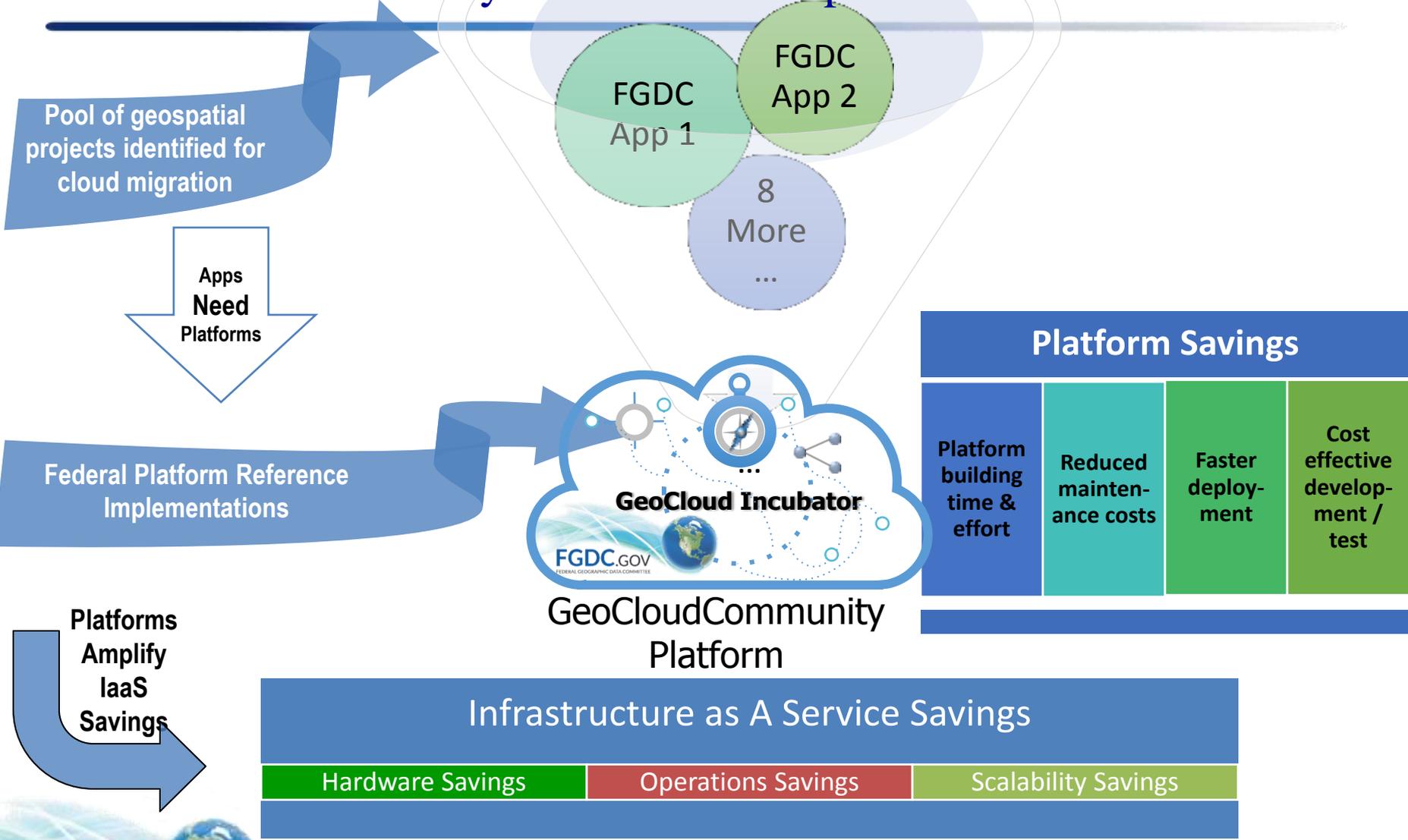
- Prototype/Pilot applications for use on a GeoCloud Platform
- Determine the application's deployment and scalability requirements
- Move to production account, operate and document metrics and experience
- Graduate to full operation as appropriate

### Objectives

- **Identify** and document the steps required to install the application
- **Document** any new application platform requirements and receive an enhanced platform from the platform development team
- **Create** and test application code bundles and installation scripts
- **Test** application for scalability and determine deployment strategy
- **Migrate** application to a production account and document



# GeoCloud Community Platform Concept and Value



Pool of geospatial projects identified for cloud migration

Apps Need Platforms

Federal Platform Reference Implementations

**GeoCloud Incubator**  
 FGDC.GOV  
 FEDERAL GEOGRAPHIC DATA COMMITTEE

GeoCloudCommunity Platform

## Platform Savings

Platform building time & effort	Reduced maintenance costs	Faster deployment	Cost effective development / test
---------------------------------	---------------------------	-------------------	-----------------------------------

Platforms Amplify IaaS Savings

## Infrastructure as A Service Savings

Hardware Savings	Operations Savings	Scalability Savings
------------------	--------------------	---------------------



# GeoCloud Sandbox History



Project Kickoff	Project Year 1 – General Purpose	Project Year 2 – Add ArcGIS	Project Year 3 – Focus on A16 Datasets	Final Project Year
<ul style="list-style-type: none"> <li>• Build manual platform</li> <li>• Develop business processes</li> <li>• Build community portal</li> <li>• Find agency candidates</li> </ul>	<ul style="list-style-type: none"> <li>• Build limited platform automation</li> <li>• Add OpenGeo to OSS platform</li> <li>• 11 projects across 6 agencies</li> </ul>	<ul style="list-style-type: none"> <li>• Increase GeoSpatial focus</li> <li>• Add GeoNode to OSS platform</li> <li>• Initial hardened ArcGIS platform</li> <li>• 15 agency projects across 10 agencies</li> </ul>	<ul style="list-style-type: none"> <li>• Revise and update OSS and ArcGIS platforms</li> <li>• Initial CKAN prototype</li> <li>• 4 projects, 4 grandfathered A16 projects</li> </ul>	<ul style="list-style-type: none"> <li>• Introduce platform.mu with full automation</li> <li>• Add Ubuntu OS</li> <li>• Migrate 2 new services – GeoNode and GeoShape</li> <li>• 6 new projects</li> </ul>

Beta Platform	Initial Automation	Add ArcGIS Platform	A16 Datasets	Full Automation
Project Kickoff	Project Year 1	Project Year 2	Project Year 3	Project Year 4
2010	2011	2012	2013	2014
				2015

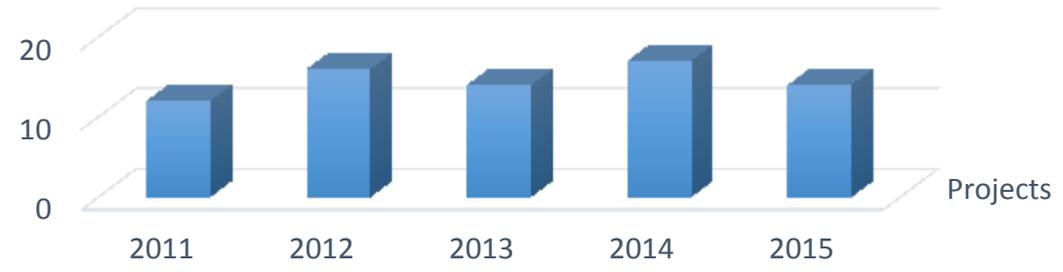
# GeoCloud Community Platform – Participating Organizations

\*Sample Participating Organizations

- Federal
- State
- Academia
- Tribal
- International



## Projects



■ Projects

# Cloud Community

## - Sample Target Applications

~37 Projects

Platform	Based On	Readiness	Base Platform	Category	Most Suitable For
<b>Wetlands Mapper</b>	Windows/IIS/ESRI	Tier 1	Windows ESRI	Normal	Public Cloud
<b>National Elevation Data Set</b>	Windows/TomCat/ORACLE /ESRI	Tier 3	Platform with additions for ORACLE as needed	Compute+ Data	Private/Community
<b>National Map</b>	Windows/TomCat/ORACLE /ESRI	Tier 3		Compute + Data	Private/Community
<b>EPA Lakes &amp; Ponds</b>	Windows / PostGres/ESRI	Tier 2			
<b>Semantic Wiki +</b>	Linux	Tier 1	Linux Open	Normal	Public Cloud
<b>NOAA Particles</b>	Linux/Java/Axis/MQ	Tier 2	Source Platform	Normal	Public Cloud
<b>GEOSS GeoNetwork w. additions</b>	Linux/TomCat/PostGres	Tier 1	with Java/Ruby on	Normal	Public Cloud
<b>TIGER Shapefiles</b>	Linux / Apache / Perl/Python	Tier 2	Rails/Tomcat/EJB3, PostGres,	Normal	Public Cloud
<b>IOOS Registry +</b>	Linux/TomCat/PostGres	Tier 2 ?	MySQL (alt)	Normal	Public Cloud
<b>ERDAP</b>	Linux/TomCat/THREDDSS	Tier 1	Python, PHP, Perl and Apache	Data	TBD
<b>Semantic Content Mgt - Drupal 7</b>	Linux LAMP	Tier 3	Good candidate for a single base platform	Normal	Public Cloud
<b>HHS Connect</b>	Linux/EJB3 (GlassFish V3)	Tier 1		Normal	Public Cloud
<b>Emergency Web Response Site</b>	Linux LAMP	Tier 3		Normal	Public Cloud
<b>OpenStreetMap</b>	Linux/MySQL/Ruby On Rails	Tier 2		Normal	Public Cloud

# GeoCloud – Pilot/Prototype Reports



National Center for Education Statistics

Amazon Hosted ESRI ArcGIS Servers  
Project Final Report

## NSDI Cooperative Agreements Program Category 7: Geospatial Platform Cloud Service Test bed Interim Report

**Date:** November 1, 2012  
**Agreement Number:** G12AC20138  
**Project title:** CA Implementation  
**Organization:** Indiana University  
 620 Union Drive, R  
 Indianapolis, IN 46  
<http://www.iupui.edu>  
**Principle Investigator:** David J Bodenham  
 V. 317-274-2455  
 F. 317-278-1830  
[infu100@iupui.edu](mailto:infu100@iupui.edu)



Figure1 August 2012 Indiana Harvest Parcels

Table 7. Cost comparison between AWS services and local server

	Cost		Amazon EC2	Local Server
Fixed (one-time) cost			None	Around \$2000 to purchase the computing instances
Monthly	Network cost		\$1 for data transferring / month	None (Included in the maintenance fee)
	Storage		\$22 /month	None (Included in the first purchase)
	Computing power	Not reserved	\$255 / month	~\$17 (Assuming the server can be used for 10 years, and then the each month cost would be (2000/(10*12)))
		Reserved	\$76 / month	
	Maintenance cost (Cooling, system, sys admin maintenance, room etc)		None	\$200 / per month (Assuming \$100 for cooling, network, and room fee, and \$100 for paying a sys admin to check and maintain the server)
Yearly	Total	Not reserved	\$3300	\$2604
		Reserved	\$1104	

Note: The cloud computing cost is based on the standard large instance with 2 CPU cores of EC2 to host GEOSS Clearinghouse. If more advanced instance types are selected, the computing power cost would increase

Figure 1: AWS Console Monitoring Interface



# GeoCloud Technical Services and Resources Tested

## GeoSpatial



## Management



## App Servers



## Platform Tools



## Operating Systems



# Examples

## – EPA How's My Waterway



### How's My Waterway

[www.epa.gov/mywaterway](http://www.epa.gov/mywaterway)

EPA's mobile web app and website, **How's My Waterway**, helps people find information on the condition of their local waterways using a smart phone, tablet, or desktop computer. It makes science-based water quality information accessible and understandable for everyone.

For decades, EPA has collected information on healthy and polluted waterways that states and territories report under the Clean Water Act. Tens of thousands of polluted waters have been identified nationally, and much has been done to reduce pollution risks to people's health, the economy and the environment. However, many Americans care the most about their local lake or stream rather than the thousands of waters described in national statistics. People want an easy way to learn about their local waters, pollution problems, why they matter, and what's being done to restore and protect them. Rather than sifting through baffling scientific information stored in complex databases, an average citizen might say, "Oh I really want to know if How's My Waterway? And please tell me in words I understand." EPA developed How's My Waterway to provide answers quickly and clearly anywhere at anytime.

#### How it Works

- SEARCH:** Use a smart phone to find out about a lake, river, or stream while standing right at the water's edge. Or check on any location in the US by entering a zip code or place name on your computer, smart phone or tablet.
- RETRIEVE:** Instantly receive a list of waterways within about five miles of the search location. Each waterway is identified as unpolluted, polluted, or unassessed, along with the year its condition was reported. A map option offers a view of the search area with the waters color-coded by assessment status. Zoom in for more details or pan across the map to check out new areas and new waters.
- DISCOVER:** Once you select a specific waterway from the map or the list of waters, the app and website offer more detailed results, including the type of pollution reported and what has been done by EPA and the states to reduce it. Technical users can follow links to detailed online scientific assessment reports.
- LEARN:** Read simple, non-technical descriptions of each type of water pollutant. These include what the pollutant is, where it comes from, how it can harm the environment, human health, or valuable economic uses of the waterway, and what you can do to help.
- FIND:** Looking for even more? The related links page connects you to popular water information on beaches, drinking water, fish habitat projects, and more!

Visit **How's My Waterway** at: [www.epa.gov/mywaterway](http://www.epa.gov/mywaterway)  
 For questions or comments, contact: <http://water.epa.gov/contactus.cfm>



# Examples – USGS BISON Tool



**USGS**  
science for a changing world

## Species Occurrence Data for the Nation

**USGS Biodiversity Information Serving Our Nation (BISON)—A Unique, Web-Based Federal Mapping Resource for Species Occurrence Data in the United States and Its Territories**

BISON's site is unprecedented, including records for most living species found in the United States (ITS, 2015) and encompassing the efforts of more than a million professional and citizen scientists. Most of BISON's species occurrence records are specific locations, not just county or state records (USGS, 2015).

Researchers benefit from a more transparent and accessible data source, which is the core of its next-generation BISON 2.0.

**Web Interface**

Users may query BISON by scientific or common name, and then refine their search results by choosing one or more criteria, including basic attributes such as observation or specimen, provider or resource status, location, higher taxa, year range, and covered subunits/divisions.

BISON also provides a refined search option for querying the database by selecting a county or state, or by drawing an exact boundary around an area of interest on the map such as protected areas, villages or even small weather areas. For instance, BISON maps more than 270,000 occurrences in New York City's Central Park alone, with detailed information available for each species record.

Species occurrence data are displayed in BISON on an interactive map (with full map or points layer options), or in checklist format. Mapped search results may be displayed on any of 30 available map layers. Users can download their search results in popular formats (CSV, Google Earth KML, or shapefile formats).

**Web Services**

In addition to the Web site, BISON has extensive Web services (http://bison.usgs.gov/) and the Application Programming Interface provides access to the BISON system for Web developers.

**Taxonomy**

BISON uses standardized scientific names for searching on the Web site and a taxonomy and synonymy applied by the Integrated Taxonomic Information System (ITIS, 2015). For example, users can search for an exact name match to see occurrences only identified as genus *Fox*, or with an ITIS-assisted search to include all taxonomic children and synonyms (and see records for all of the species, subspecies, and varieties of *Fox*, along with all of their synonyms). Names not in ITIS can be searched only as exact matches, but 95 percent of the taxa represented in BISON are listed in ITIS. BISON also offers an ITIS Enabled Common Name Search option.

**National and International Partnerships**

In agreement with the National Science Foundation (NSF), BISON serves as the Federal component to iDigBio (iDigBio, 2015), which is part of the NSF-funded Advancing Digitization of Biodiversity Collections (ADBC) program (Hammer, 2013). BISON is also the Biodiversity Hub of the multi-agency Foundation-based Open Resources and Machine Accessibility (GeoINFORMA) initiative proposed by the President's Council.

**USGS Department of the Interior  
U.S. Department of Energy**

**Full Report 2015-2016  
September 2015**

**USGS**  
science for a changing world

**Biodiversity Information Serving Our Nation (BISON) - U.S. species occurrence data & maps**

Home About Data Providers Statistics API Examples Blog Help

ITIS Enabled Search by Scientific Name

Search Reset

Refine Your Search Previous Search (0) 261,702,042 results for all species using ITIS taxonomy

Points Layer County Heatmap State Heatmap Map Checklist

Legend

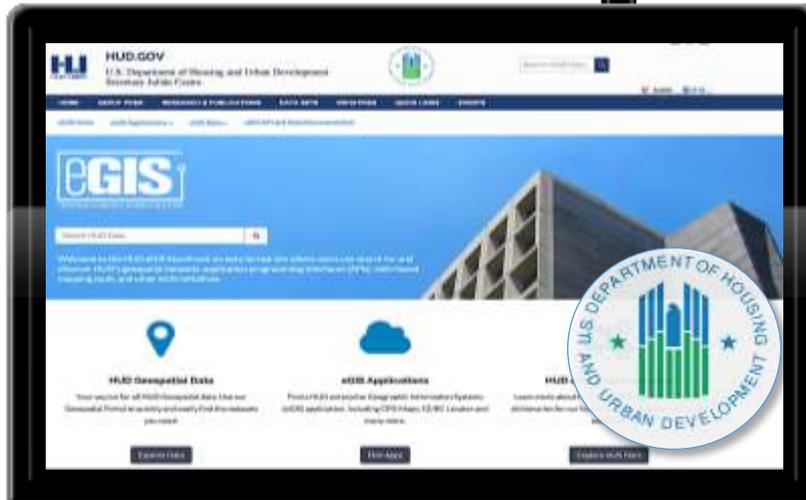


# Examples

## – HUD and EPA



NEPAnode provides a suite of geospatial enriched shared web services, software, and GIS tools to assist practitioners, resource managers and decision makers in reporting and understanding environmental assessments and permitting.



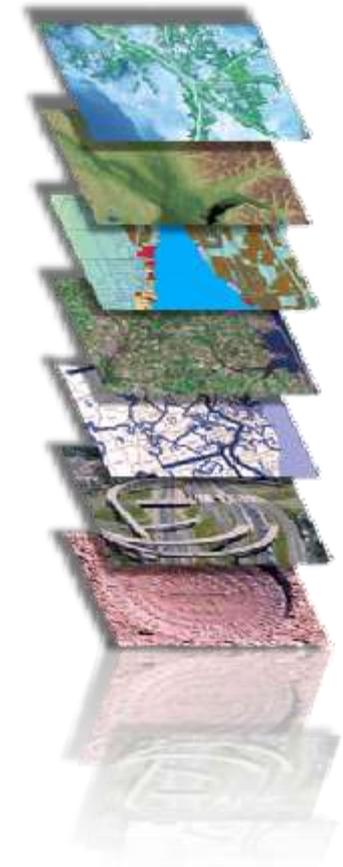
The HUD Storefront unifies public access to HUD's portfolio of applications and geospatial data resources, including the 11 National Geospatial Data Asset (NGDA) datasets that HUD is currently contributing to the Federal geospatial data portfolio.

# GeoCloud

## – Lessons Learned

---

- **Collaboration:** Community of Practice was a positive method for this effort
- **Cost Efficiency:** Cloud offers strong potential for operational savings
- **Architecture:** Application re-engineering generally necessary to take optimize use
- **Security:** Securing Cloud Services very challenging



# GeoCloud

## – Recommendations

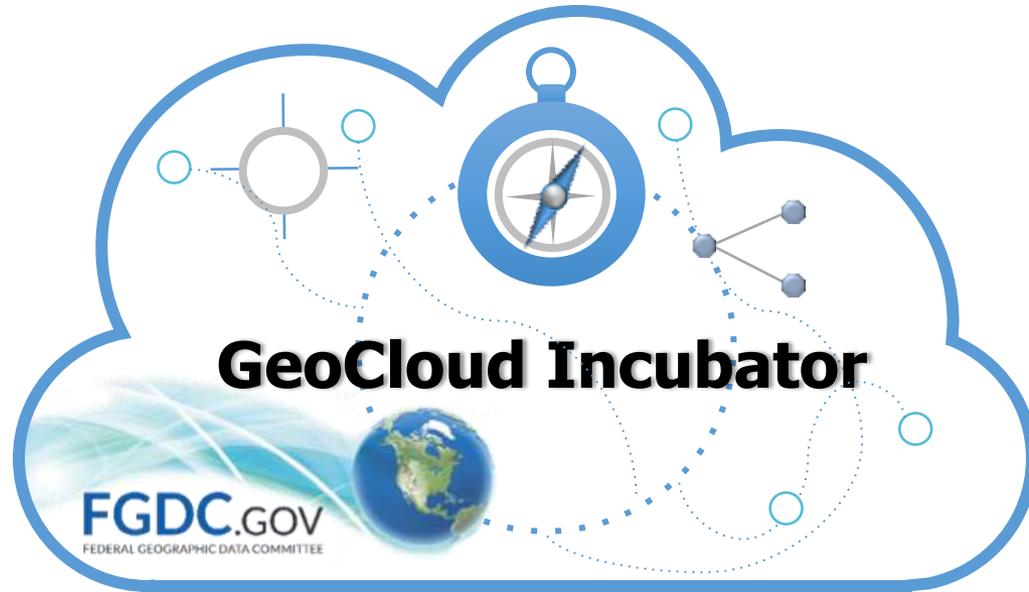
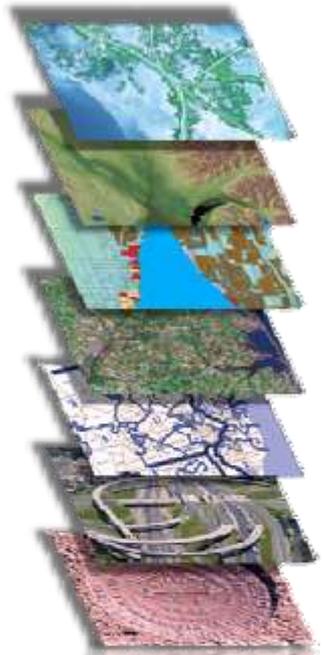
---



- Continue Collaborative R&D of Emerging Cloud Services
- Education and Awareness
- Share Lessons Learned
- Establish Resources via the GeoPlatform
- Security Architecture Development and Efficiency

# Summary

---



Questions?

Rich Frazier  
efrazier@usgs.gov

