

Data Management, Discovery and Access Technology for Natural Science and Beyond

Karl Benedict

Geospatial and Environmental Data Scientist, CUL&LS
Director, EDAC
kbene@unm.edu

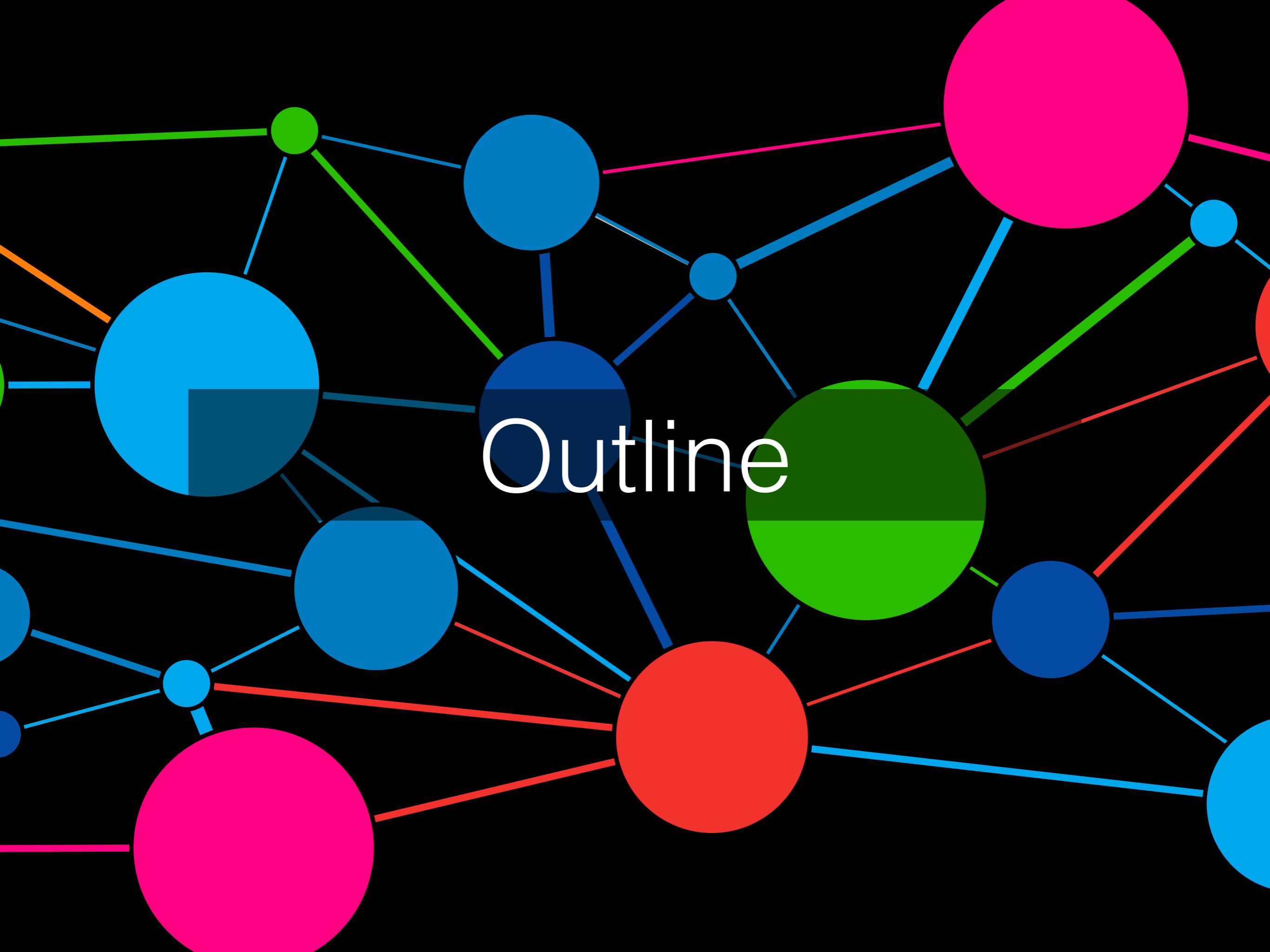
<http://tinyurl.com/ksvpgse>





Outline

Outline

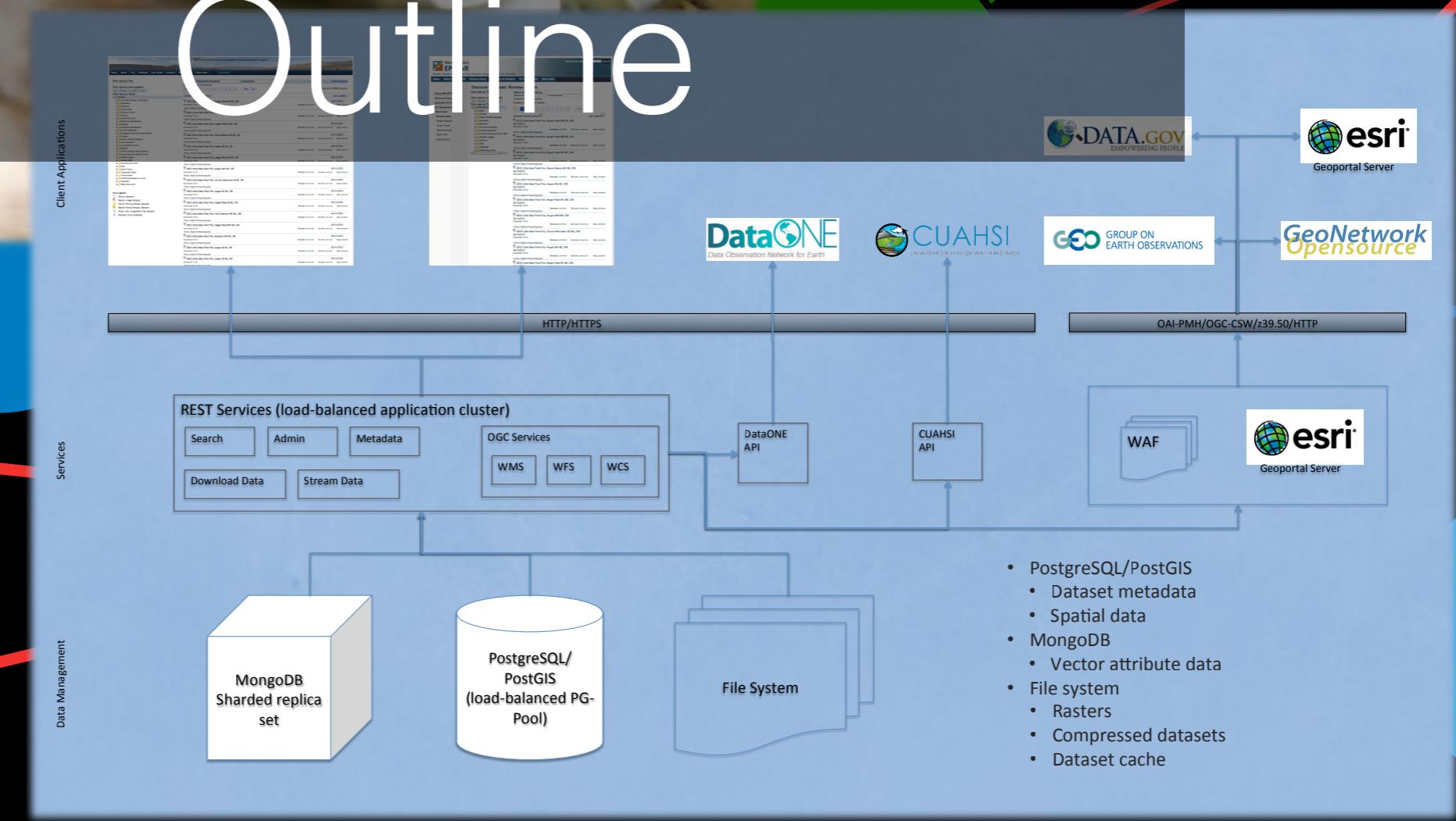




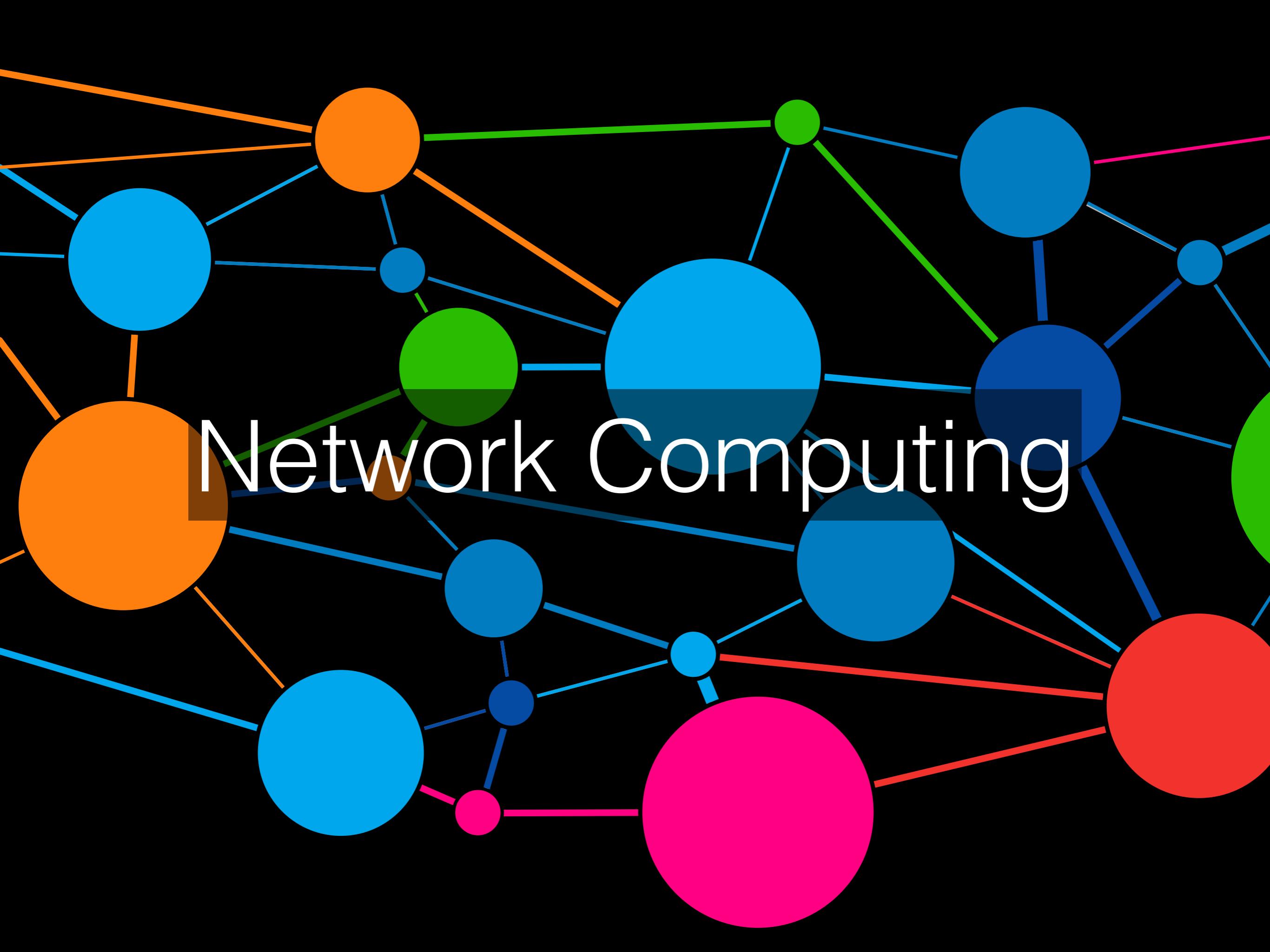
Outline



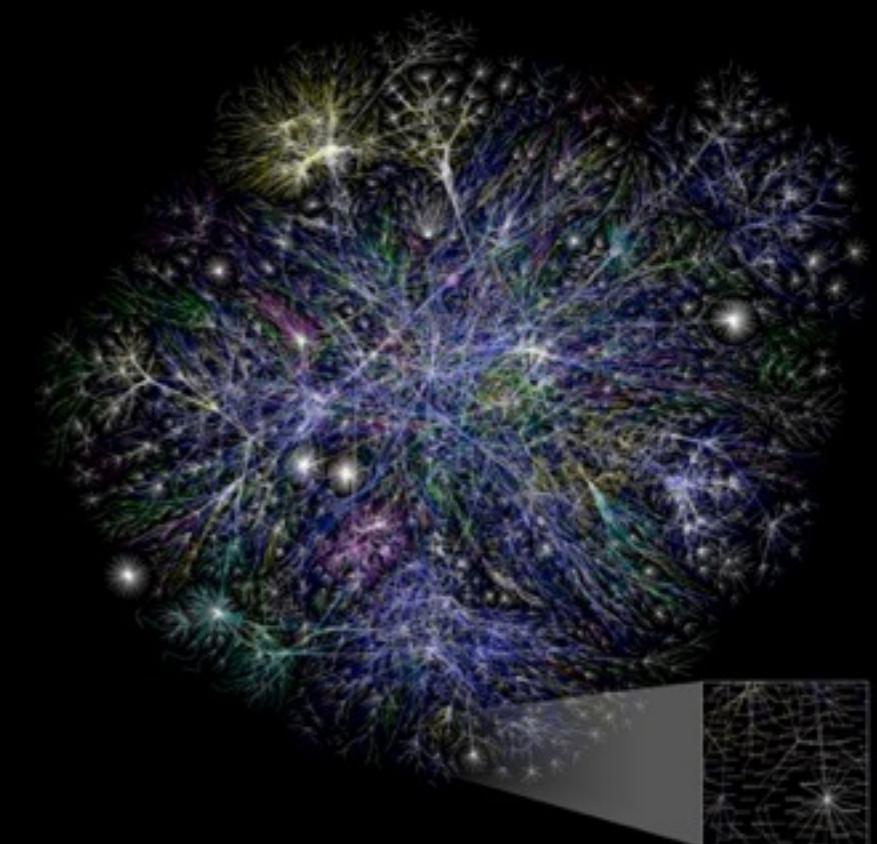
Outline



Network Computing

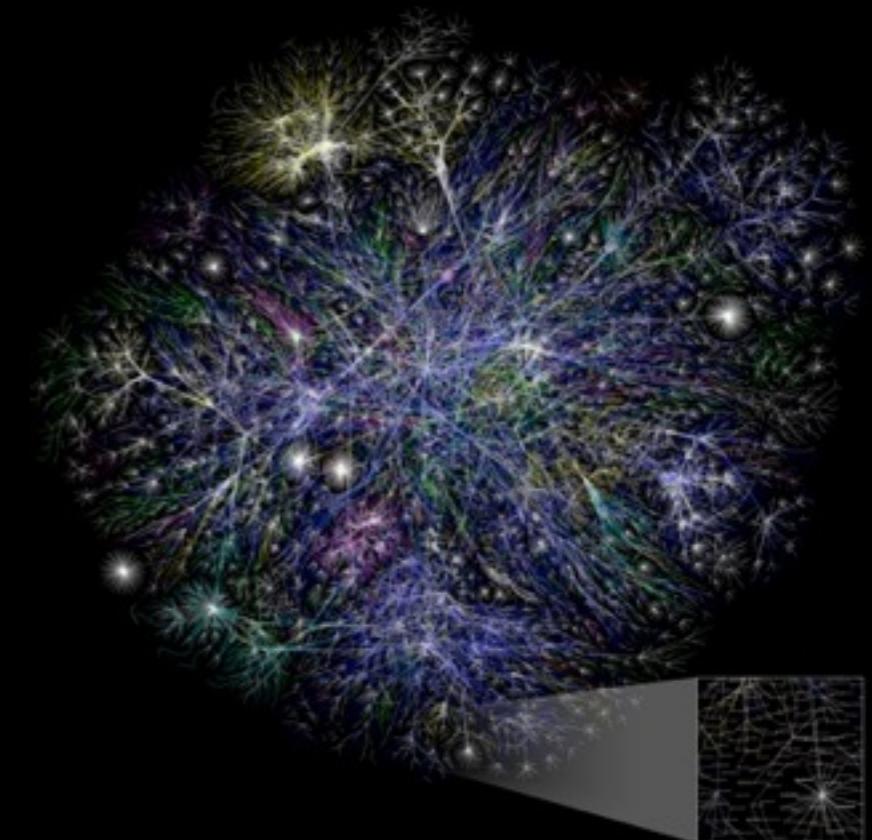


Network Computing



Network Computing

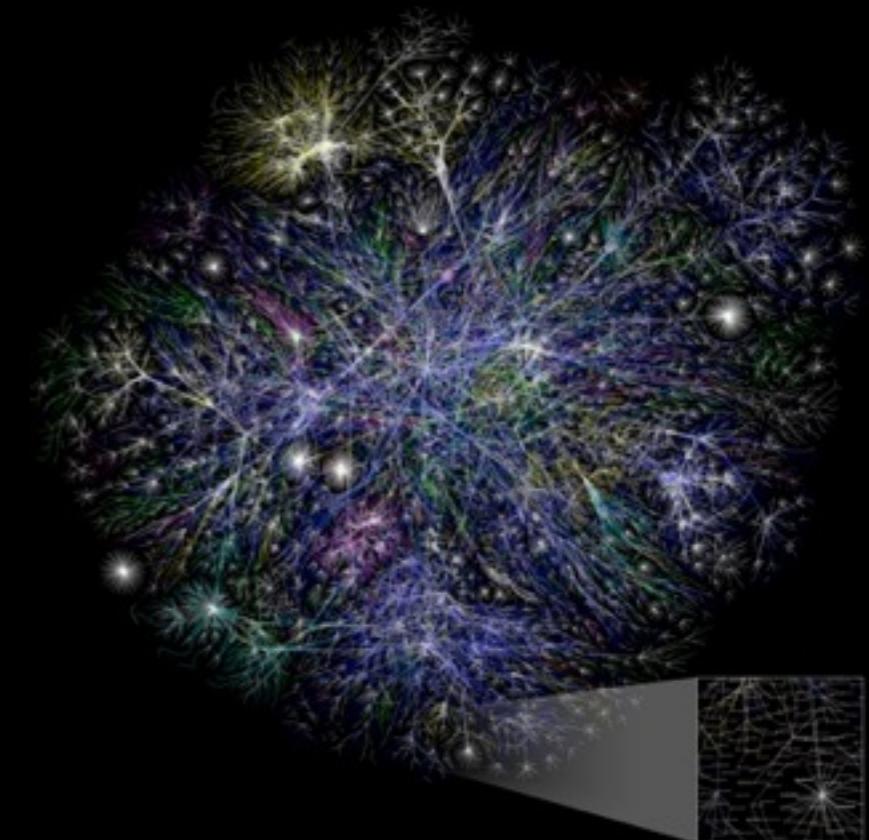
- Tremendous interconnectivity of computers via the World Wide Web and the Internet



Partial Internet Map (circa 2005) - Matt Britt

Network Computing

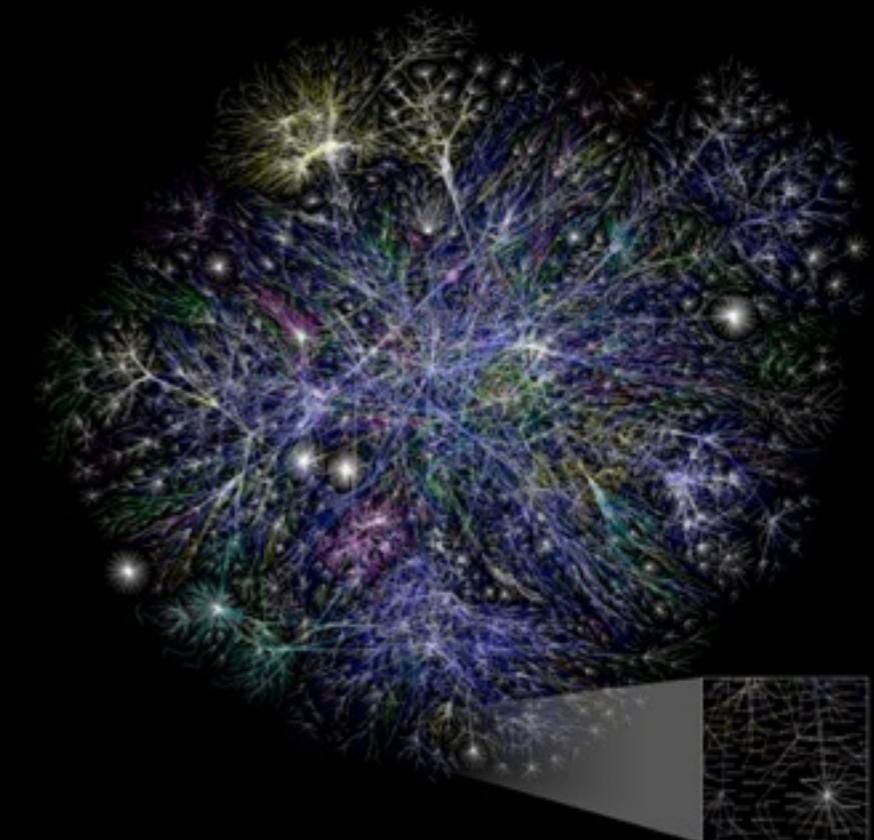
- Tremendous interconnectivity of computers via the World Wide Web and the Internet
- Emergence of Software as a Service (SaaS)



Partial Internet Map (circa 2005) - Matt Britt

Network Computing

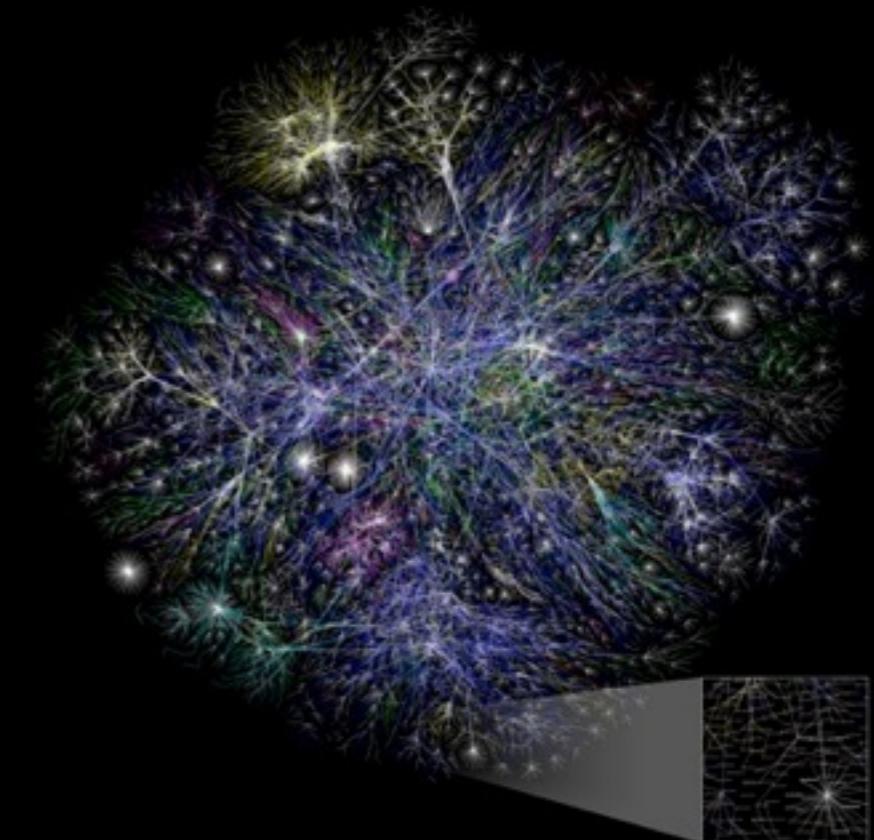
- Tremendous interconnectivity of computers via the World Wide Web and the Internet
- Emergence of Software as a Service (SaaS)
- Web application “mashups” enabled by the publication of Application Programming Interfaces (API)
 - Google Maps as an example



Partial Internet Map (circa 2005) - Matt Britt

Network Computing

- Tremendous interconnectivity of computers via the World Wide Web and the Internet
- Emergence of Software as a Service (SaaS)
- Web application “mashups” enabled by the publication of Application Programming Interfaces (API)
 - Google Maps as an example
- Emergence of interoperability standards that enable combination of content from multiple Internet resources



Partial Internet Map (circa 2005) - Matt Britt

In a Phrase ...

The current networking
computing model consists of
Components Interacting with
Each Other

In a Phrase ...

The current networking
computing model consists of
Components Interacting with
Each Other

In a Phrase ...

The current networking
computing model consists of
Components Interacting with
Each Other

So - we Need to work on
answering the following Questions

So - we Need to work on
answering the following Questions

- What are components?

So - we Need to work on
answering the following Questions

- What are components?
- What does it mean to interact?

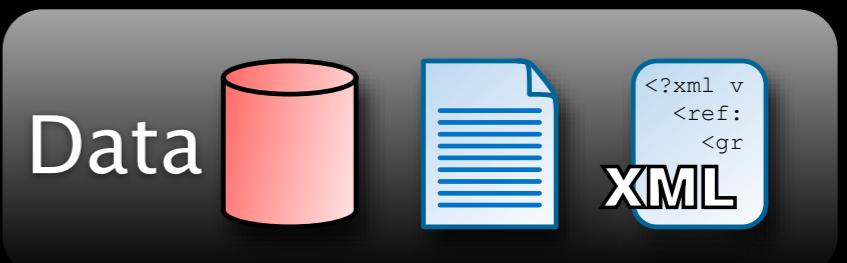
The Big Picture - Services Oriented Architectures

The Big Picture - Services Oriented Architectures

- Services Oriented Architecture (SOA) for Geospatial Data and Processing

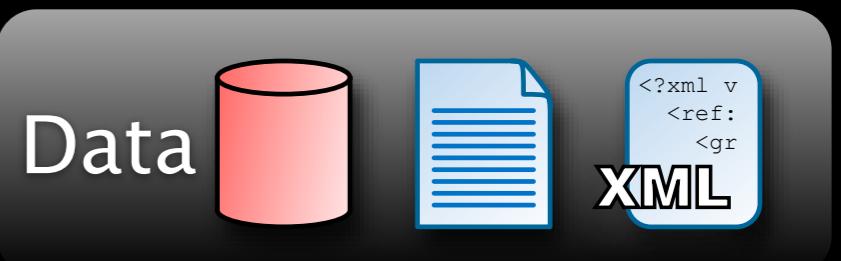
The Big Picture - Services Oriented Architectures

- Services Oriented Architecture (SOA) for Geospatial Data and Processing
 - Data, Processing, Client Tiers



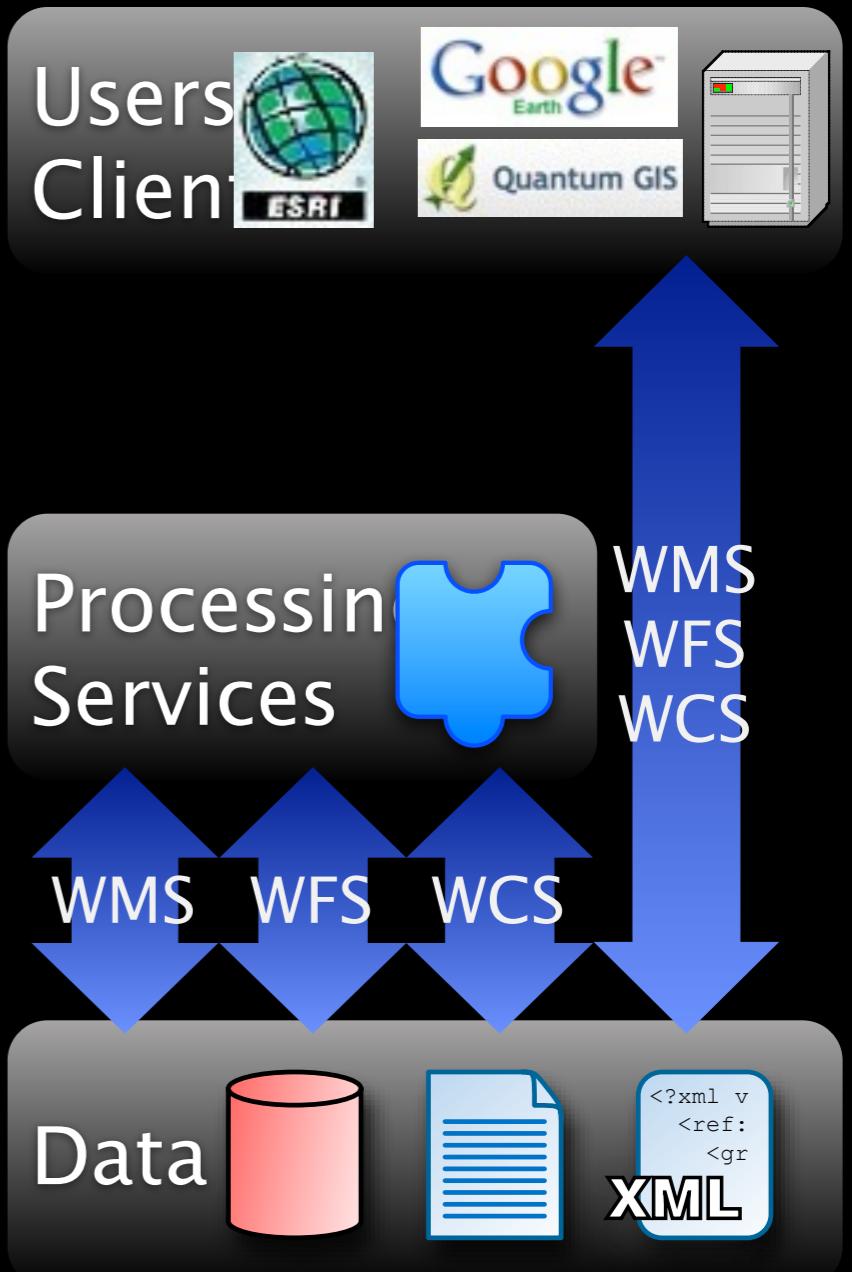
The Big Picture - Services Oriented Architectures

- Services Oriented Architecture (SOA) for Geospatial Data and Processing
 - Data, Processing, Client Tiers
- Open Geospatial Consortium Interoperability Standards



The Big Picture - Services Oriented Architectures

- Services Oriented Architecture (SOA) for Geospatial Data and Processing
 - Data, Processing, Client Tiers
- Open Geospatial Consortium Interoperability Standards
 - WMS, WFS, WCS



The Big Picture - Services Oriented Architectures

- Services Oriented Architecture (SOA) for Geospatial Data and Processing
 - Data, Processing, Client Tiers
- Open Geospatial Consortium Interoperability Standards
 - WMS, WFS, WCS
- Geospatial Metadata Standards
 - ISO 19115, FGDC



The Big Picture - Services Oriented Architectures

- Services Oriented Architecture (SOA) for Geospatial Data and Processing
 - Data, Processing, Client Tiers
- Open Geospatial Consortium Interoperability Standards
 - WMS, WFS, WCS
- Geospatial Metadata Standards
 - ISO 19115, FGDC
- Internet Standards
 - Web: HTML, CSS, JavaScript, XML, RDF



The Big Picture - Services Oriented Architectures

- Services Oriented Architecture (SOA) for Geospatial Data and Processing
 - Data, Processing, Client Tiers
- Open Geospatial Consortium Interoperability Standards
 - WMS, WFS, WCS
- Geospatial Metadata Standards
 - ISO 19115, FGDC
- Internet Standards
 - Web: HTML, CSS, JavaScript, XML, RDF
 - SOAP - Simple Object Access Protocol
 - REST - Representation State Transformation



The Big Picture - Services Oriented Architectures

- Services Oriented Architecture (SOA) for Geospatial Data and Processing
 - Data, Processing, Client Tiers
- Open Geospatial Consortium Interoperability Standards
 - WMS, WFS, WCS
- Geospatial Metadata Standards
 - ISO 19115, FGDC
- Internet Standards
 - Web: HTML, CSS, JavaScript, XML, RDF
 - SOAP - Simple Object Access Protocol
 - REST - Representation State Transformation
- Proprietary Protocols



The Big Picture - Services Oriented Architectures

- Services Oriented Architecture (SOA) for Geospatial Data and Processing
 - Data, Processing, Client Tiers
- Open Geospatial Consortium Interoperability Standards
 - WMS, WFS, WCS
- Geospatial Metadata Standards
 - ISO 19115, FGDC
- Internet Standards
 - Web: HTML, CSS, JavaScript, XML
 - SOAP - Simple Object Access Protocol
 - REST - Representation State Transformation
- Proprietary Protocols



The Big Picture - Services Oriented Architectures

- Services Oriented Architecture (SOA) for Geospatial Data and Processing
 - Data, Processing, Client Tiers
- Open Geospatial Consortium Interoperability Standards
 - WMS, WFS, WCS
- Geospatial Metadata Standards
 - ISO 19115, FGDC
- Internet Standards
 - Web: HTML, CSS, JavaScript, XML
 - SOAP - Simple Object Access Protocol
 - REST - Representation State Transformation
- Proprietary Protocols



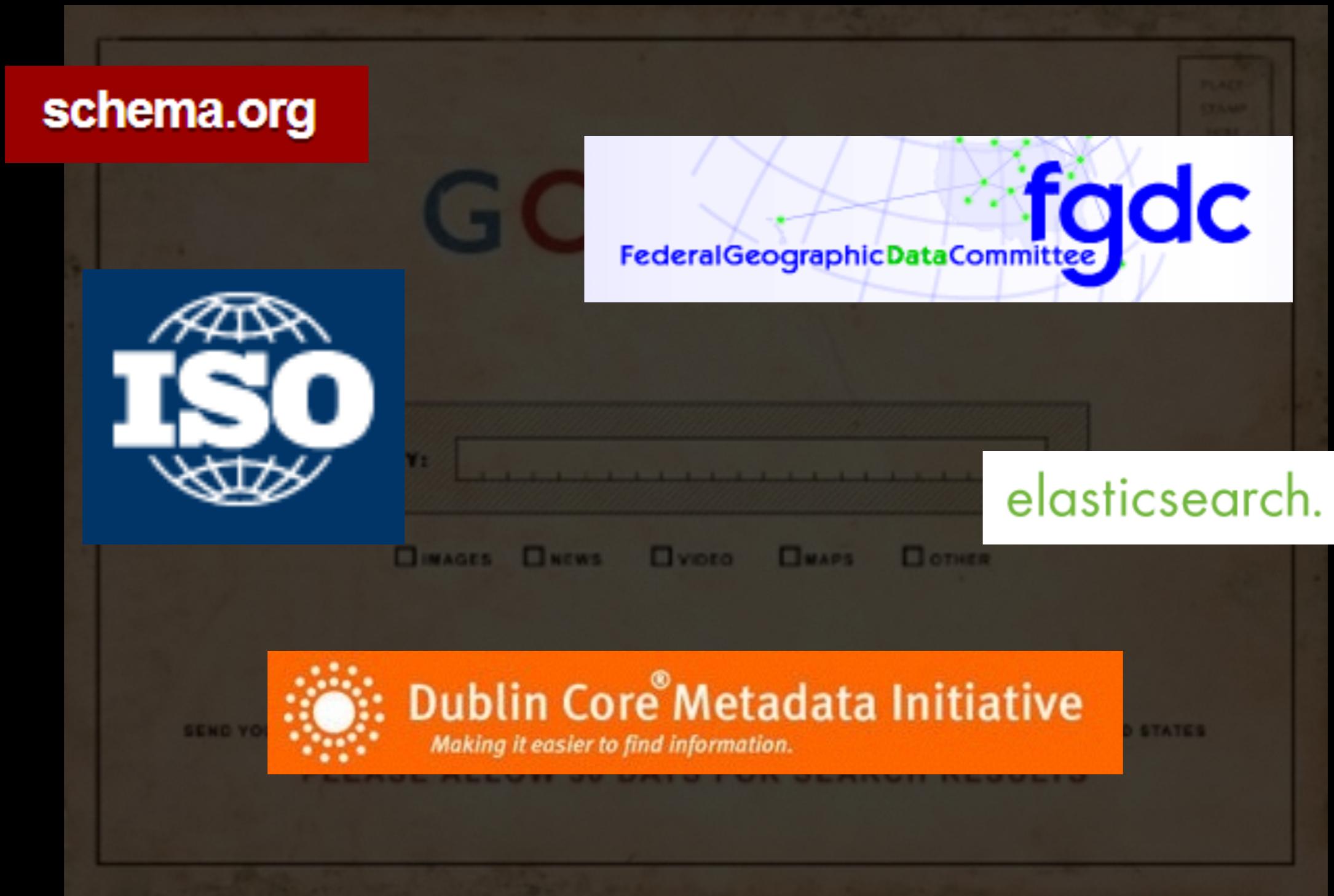


Discovery & Access

Discovery - Google for Data?



Discovery - Google for Data?



Playing With Others



The screenshot shows the Data.gov homepage with a blue header bar. On the left is the Data.gov logo with an American flag icon. The top navigation bar includes links for DATA, TOPICS, IMPACT, APPLICATIONS, DEVELOPERS, and CONTACT. A search bar with a magnifying glass icon is at the top right. Below the header is a large blue banner with the text "DATA CATALOG". Underneath is a search interface with a "Datasets" dropdown, "Organizations" and "Interactive Datasets" buttons, and a search input field containing "Earth Data Analysis Center" with a search icon. To the left is a "Filter by location" section with a map of the world and a "Clear" button. At the bottom are buttons for "Dataset Type" (AJ, 14), and "File Format" (CSV, XML, HTML).



Access - As Simple as a Web Link

The screenshot shows the RGIS website interface. At the top, there's a navigation bar with links for Home, About, FAQ, Feedback, User Guide, Glossary, Web Services, Order Help, and Get Data. Below the navigation is a search bar with fields for 'Define Area of Interest by Planname' and 'or Quadrangle'. A 'Current theme' dropdown is set to 'Climate - SNOTEL'. There are also buttons for 'Advanced Search', 'Filter data by Title', 'Filter data by when updated', and 'Filter data by Theme'. The 'Theme' filter is expanded, showing categories like AEROS, Boundaries, Cultural, Census Data, Cities and Towns, Climate, Digital OrthophOTOGRAPHY, Elevation, Emergency Management, Geographic Name and Feature Names, Geology, Historic Aerial Photography, Land Ownership, Land Use and Cover, Links, MNDOT, PLSS (Township, Range, Section), Quad Grids and Geodetic Control, Satellite Imagery, Shaded Relief, Socioeconomic Data, Soil, State Forestry, Topographic Maps, Transportation, USGS Quadrangles by County, Vegetation, and Water Resources. The main content area displays a list of SNOTEL datasets, each with a thumbnail, title, last update date (e.g., 04/02/2014), metadata XML link, services link, and a map preview link. The datasets listed include SNOTEL Raton, NM (Mar, 2014); SNOTEL Chamita, NM (Mar, 2014); SNOTEL Elk Cabin, NM (Mar, 2014); SNOTEL Frisco Divide, NM (Mar, 2014); SNOTEL Gallegos Peak, NM (Mar, 2014); SNOTEL Hopewell, NM (Mar, 2014); SNOTEL Lookout Mountain, NM (Mar, 2014); and SNOTEL McKnight Cabin, NM (Mar, 2014). At the bottom right of the content area, there are links for 'GSToRE V3', 'Stable', 'Experimental', and 'Architecture'.

<http://rgis.unm.edu>

<http://gstore.unm.edu/docs/index.html>

GSToRE V3 API Documentation

Geographic Storage, Transformation and Retrieval Engine Version 3

Architecture

A brief description of the GSToRE architecture.

[View docs](#)

Stable API

Versioned, stable services.

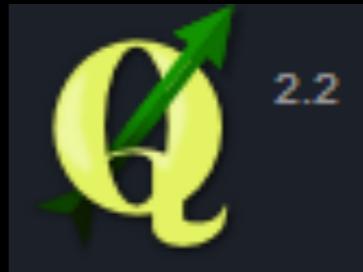
[View docs](#)

Experimental API

Services listed as experimental may undergo significant changes to the routing or service responses. Use at your own risk.

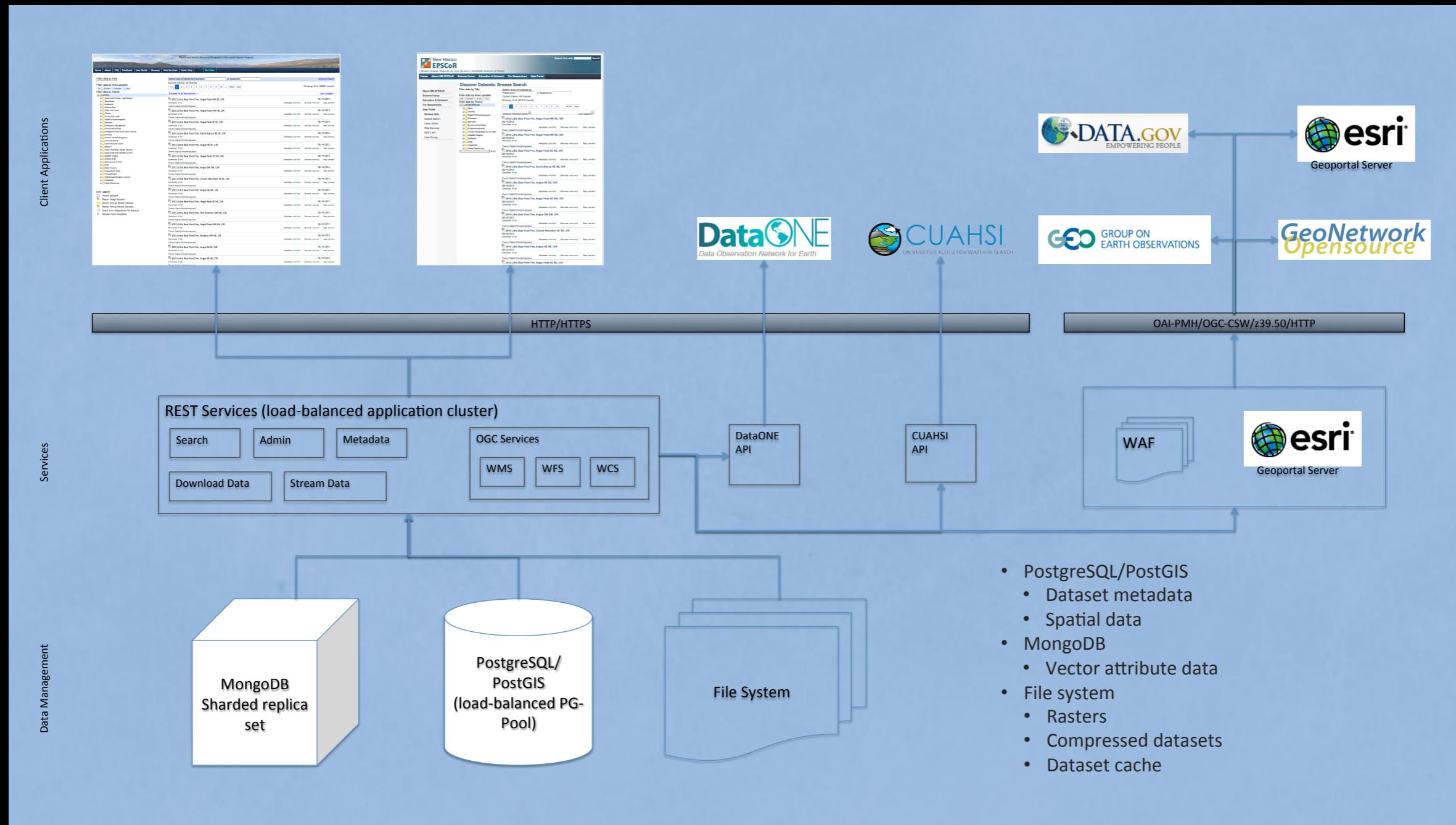
[View docs](#)

... even for existing
analysis tools



GSToRE Platform

Introduction

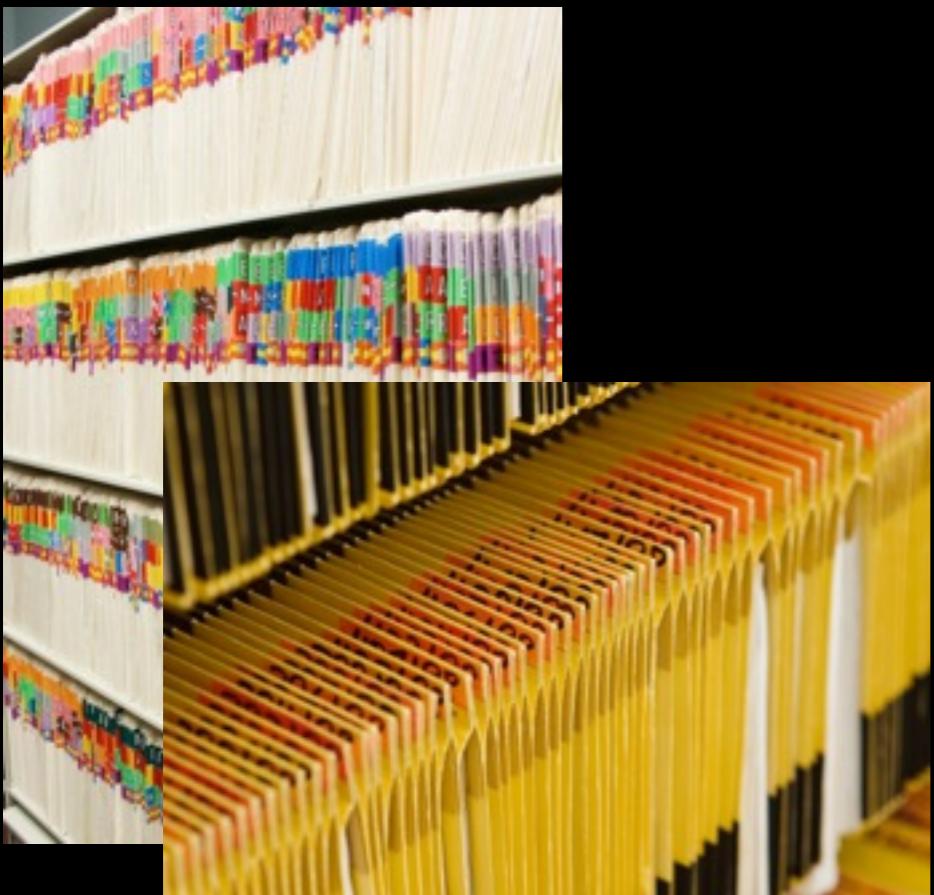


Big Data

Many Problems and Solutions*

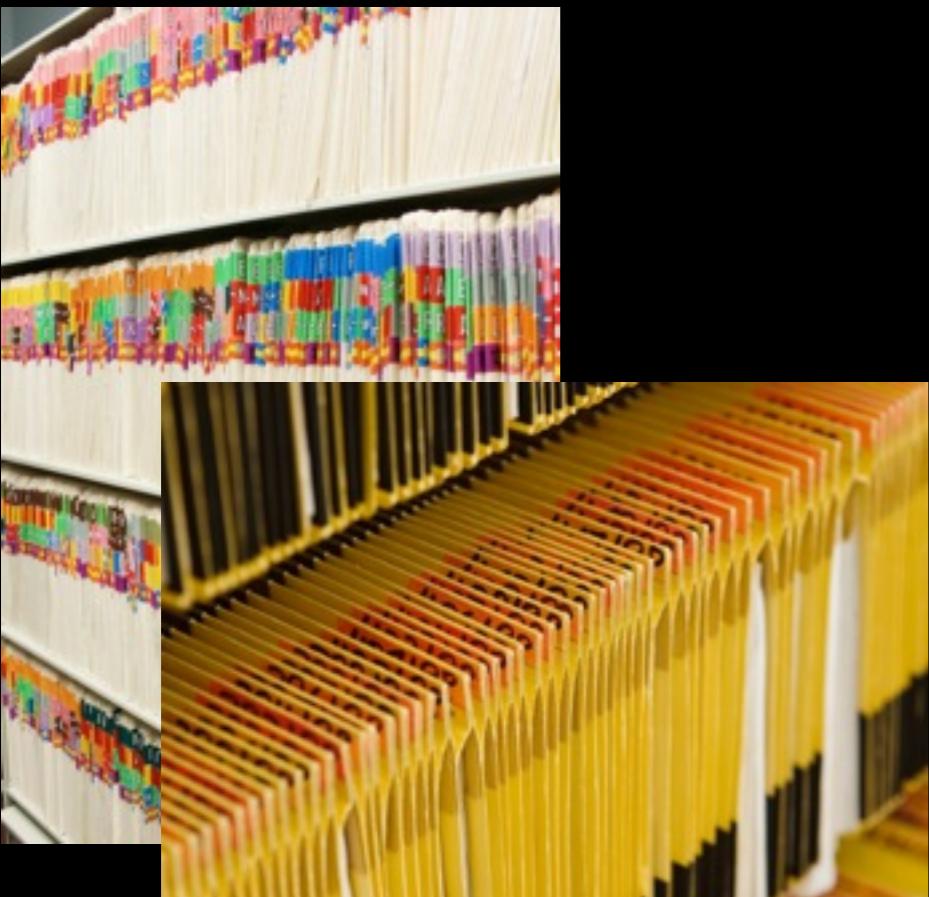
Big Data

Many Problems and Solutions*



Big Data

Many Problems and Solutions*



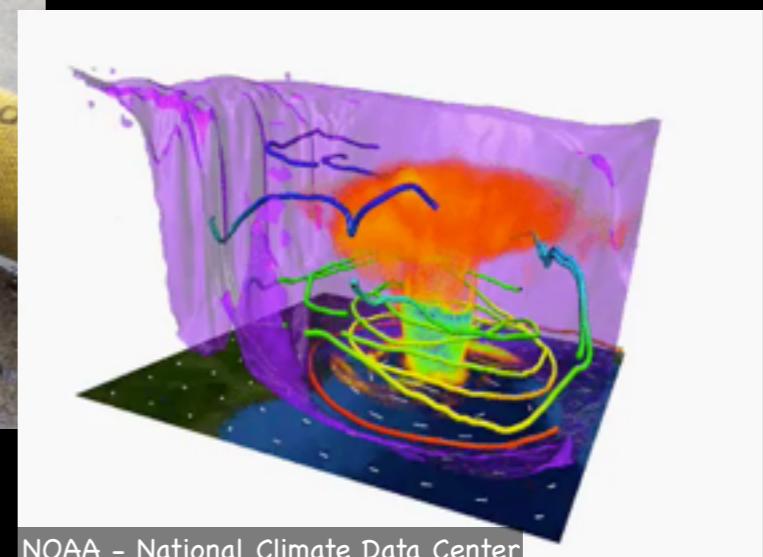
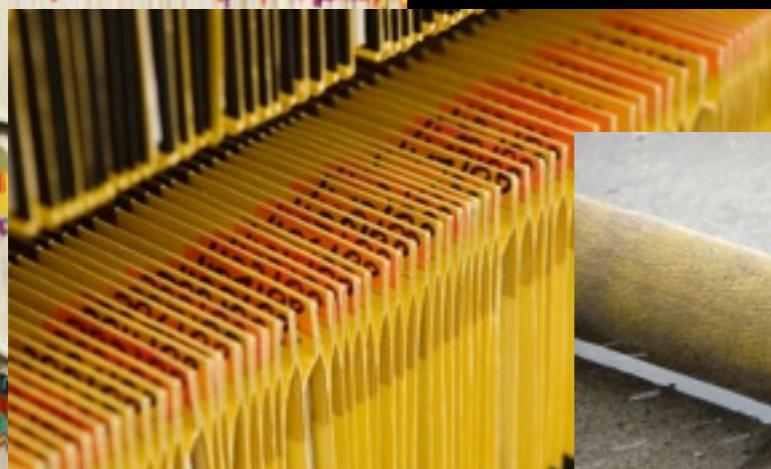
Big Data

Many Problems and Solutions*



Big Data

Many Problems and Solutions*



NOAA - National Climate Data Center

Big Data Sample Solutions

- Horizontal vs. vertical scaling
- Unstructured and semi-structured data models
- Parallelism
- Linked analytics
- Machine learning

Big Data

An Aligned Definition

Big Data

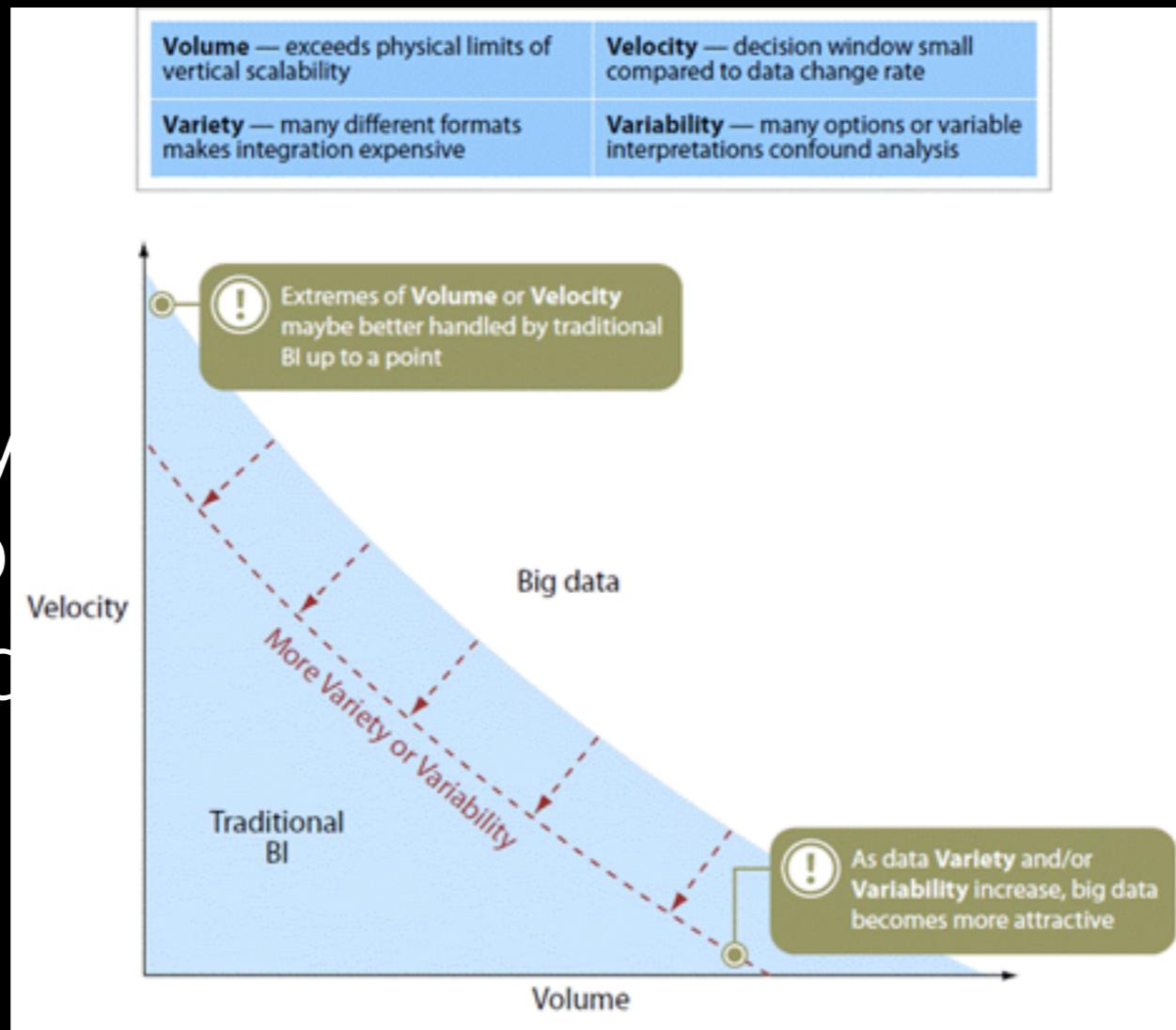
An Aligned Definition

“An easily scalable system of unstructured data with accompanying tools that can efficiently pull structured datasets.”*

Big Data

An Aligned Definition

“An easily
with acco
structured



http://blogs.forrester.com/brian_hopkins/11-08-29-big_data_brewer_and_a_couple_of_webinars

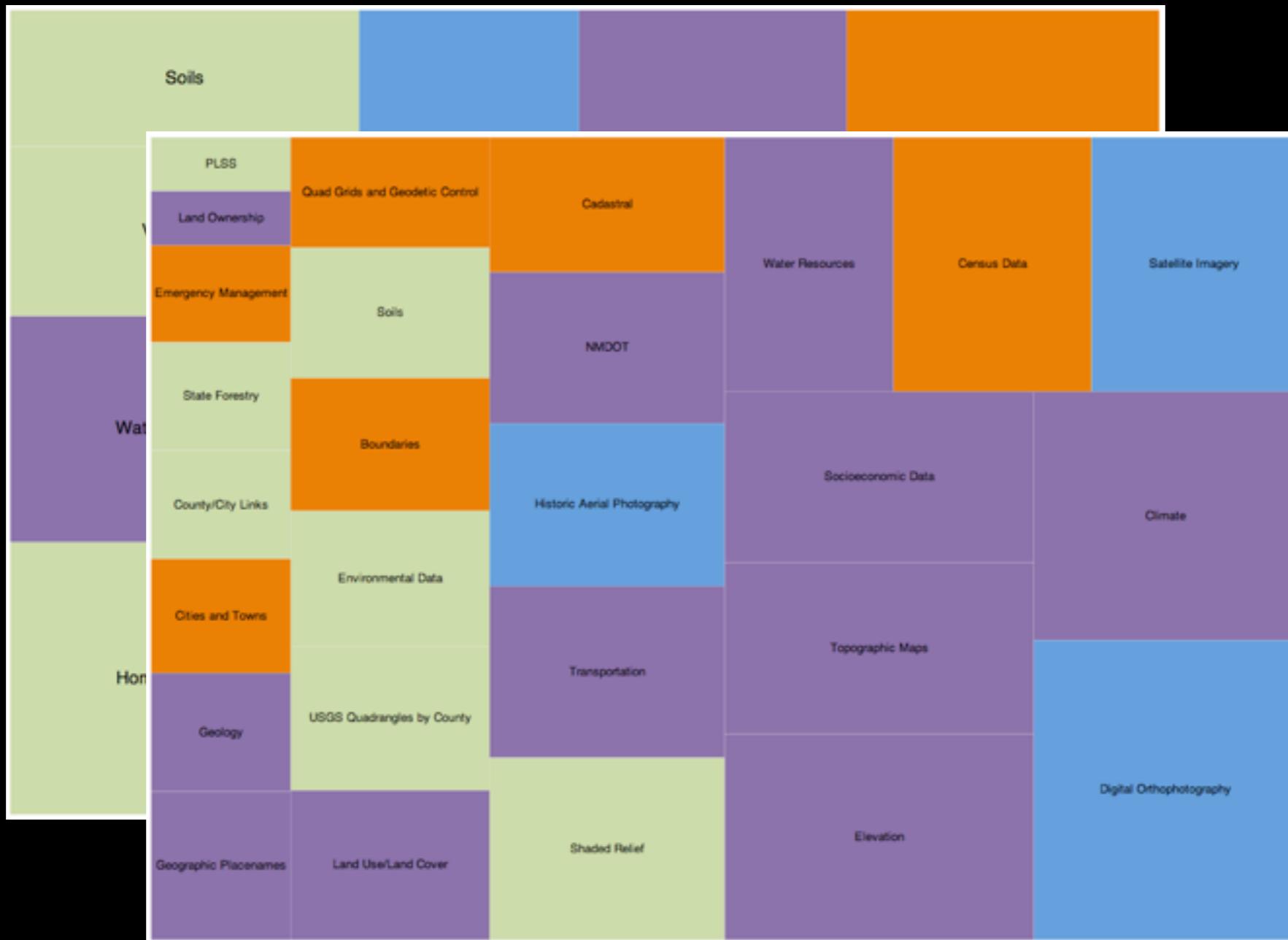
tured data
ciently pull

EDAC's Data Challenge

A Snapshot

EDAC's Data Challenge

A Snapshot



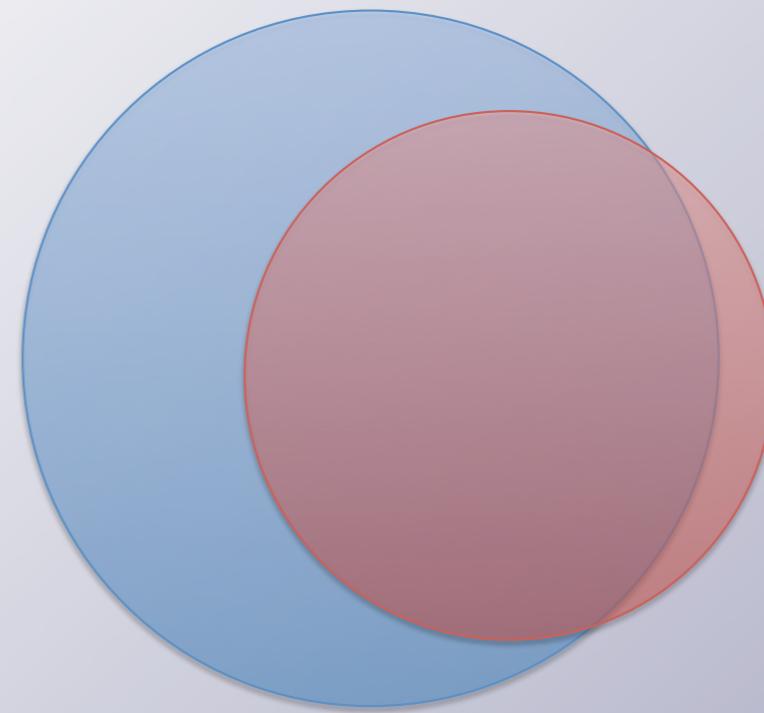
EDAC's Data Challenge

A Snapshot



NM EPSCoR

- 281,315 datasets
 - 9,054 files
 - 169,981 vectors
 - 102,280 rasters



RGIS

- 120,031 datasets
 - 12,171 files
 - 11,808 vectors
 - 96,052 rasters

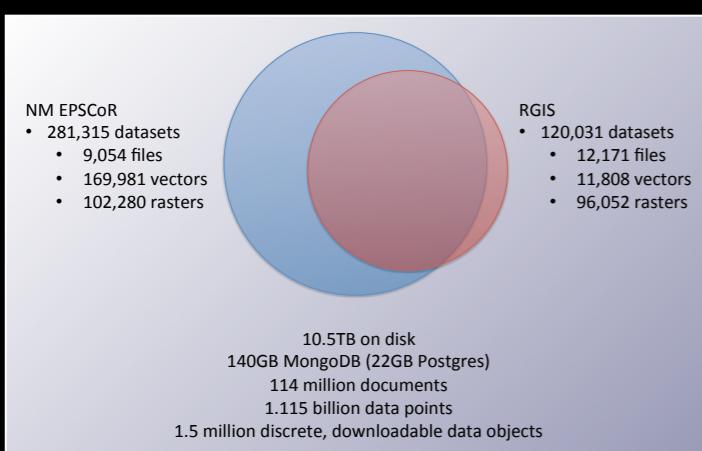
10.5TB on disk
140GB MongoDB (22GB Postgres)
114 million documents
1.115 billion data points
1.5 million discrete, downloadable data objects

EDAC's Data Challenge

A Snapshot

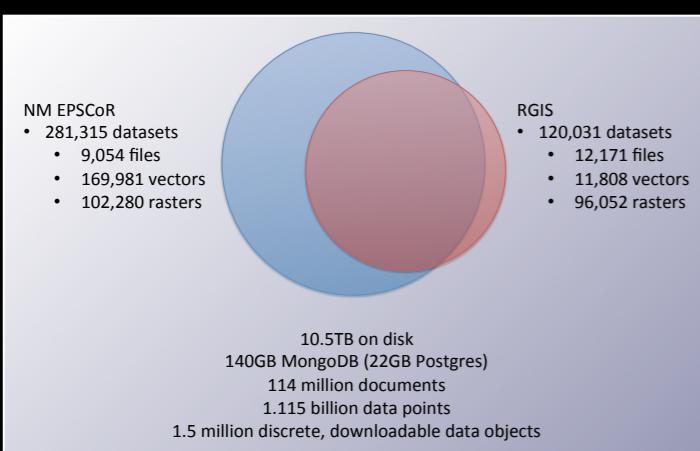


Metadata	Vectors	Rasters	Files	Services
FGDC	SHP	GeoTIFF	ZIP	WMS
FGDC-RSE	KML	IMG	HTML	WFS
ISO 19115-2 / 19139	GML	SID	PDF	WCS
ISO 19119	GeoJSON	ECW	DOC/DOCX	
ISO 19110	JSON	DEM	GZ	
	CSV	ASCII	XLS/XLSX	
	XLS		PPT/PPTX	



EDAC's Data Challenge

A Snapshot



Metadata	Vectors	Rasters	Files	Services
FGDC	SHP	GeoTIFF	ZIP	WMS
FGDC-RSE	KML	IMG	HTML	WFS
ISO 19115-2 / 19139	GML	SID	PDF	WCS
ISO 19119	GeoJSON	ECW	DOC/DOCX	
ISO 19110	JSON	DEM	GZ	
	CSV	ASCII	XLS/XLSX	
	XLS		PPT/PPTX	

EDAC's Data Challenge

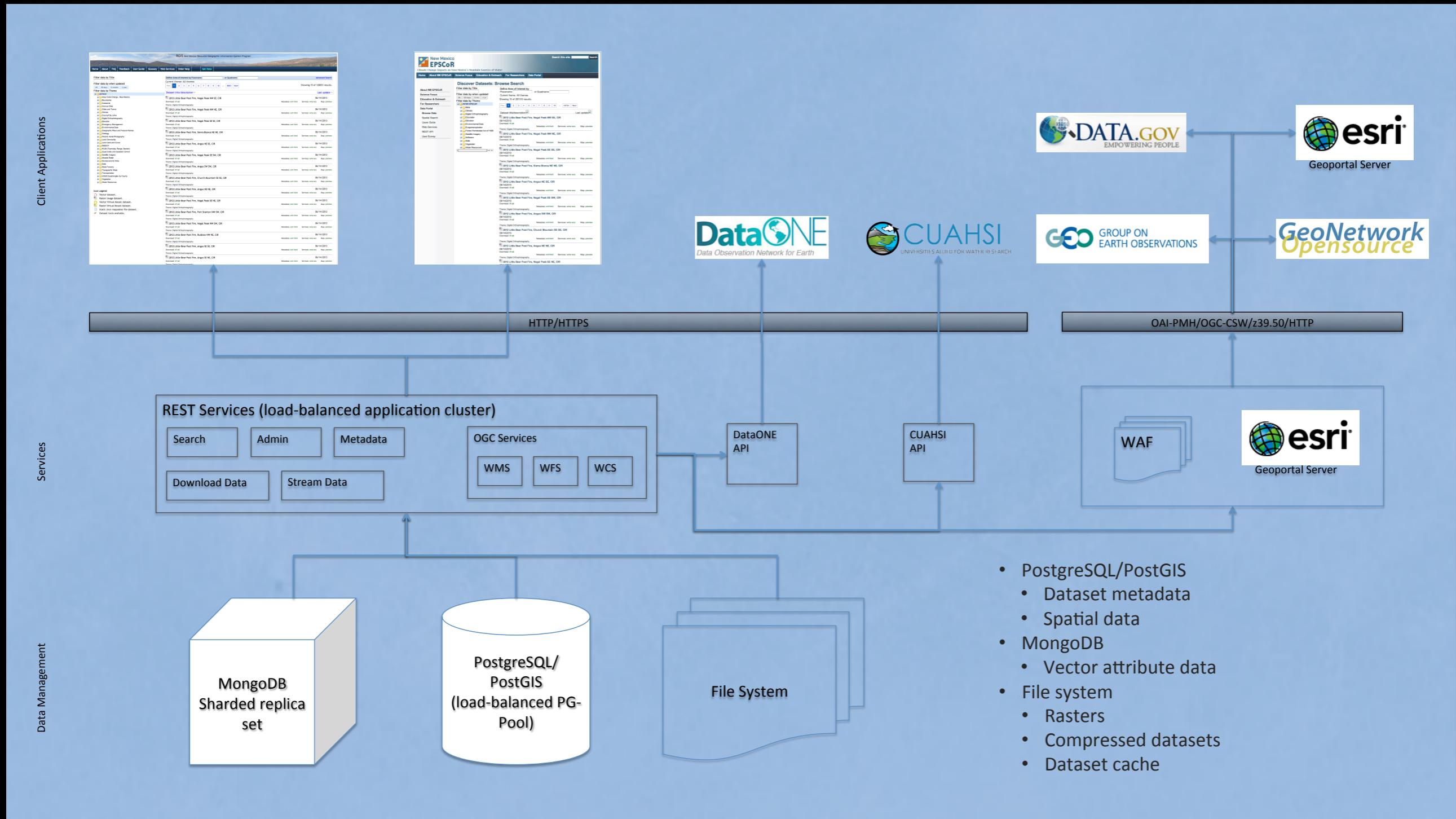
A Snapshot

EDAC's Data Challenge

Our Solution

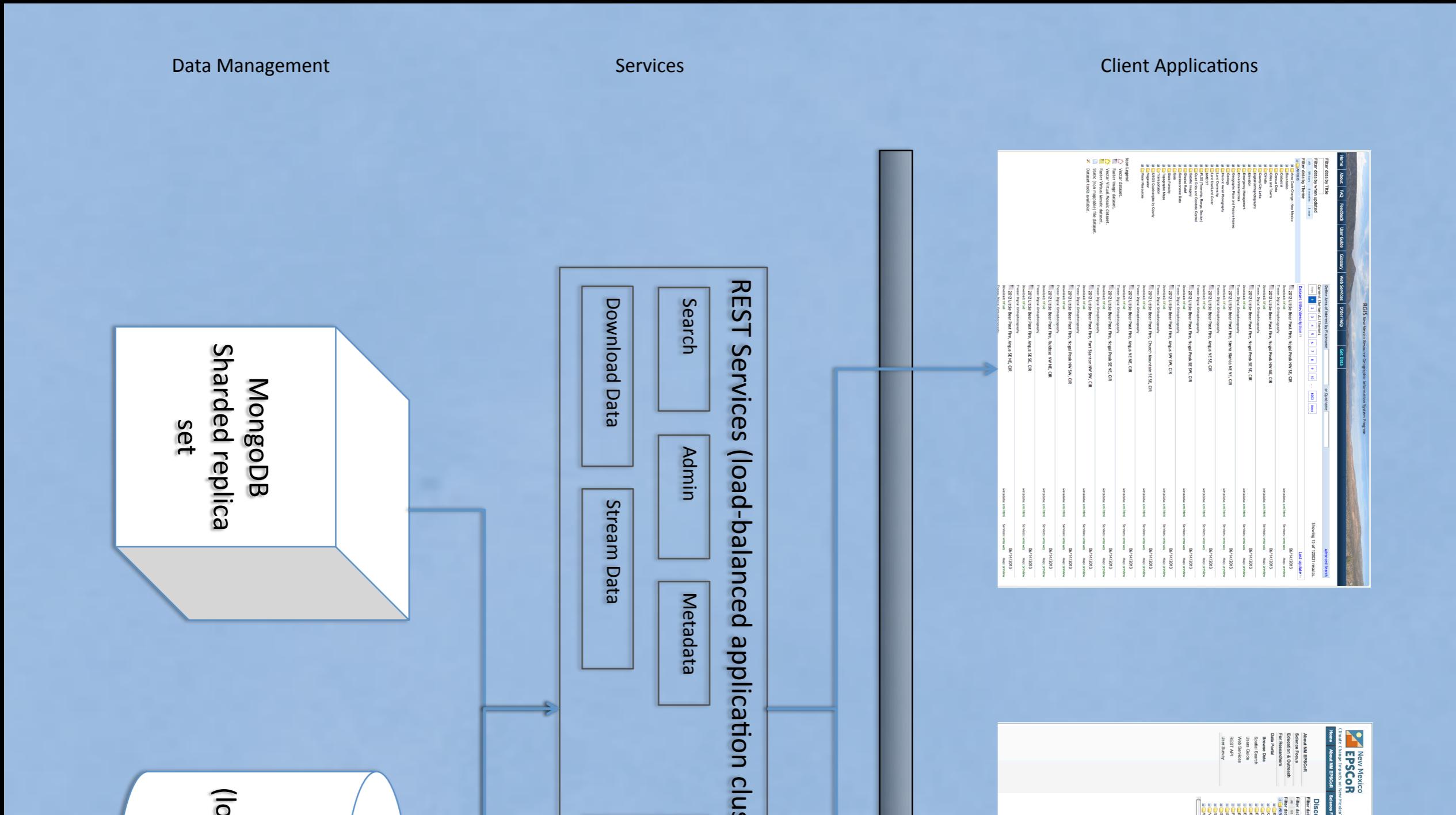
EDAC's Data Challenge

Our Solution



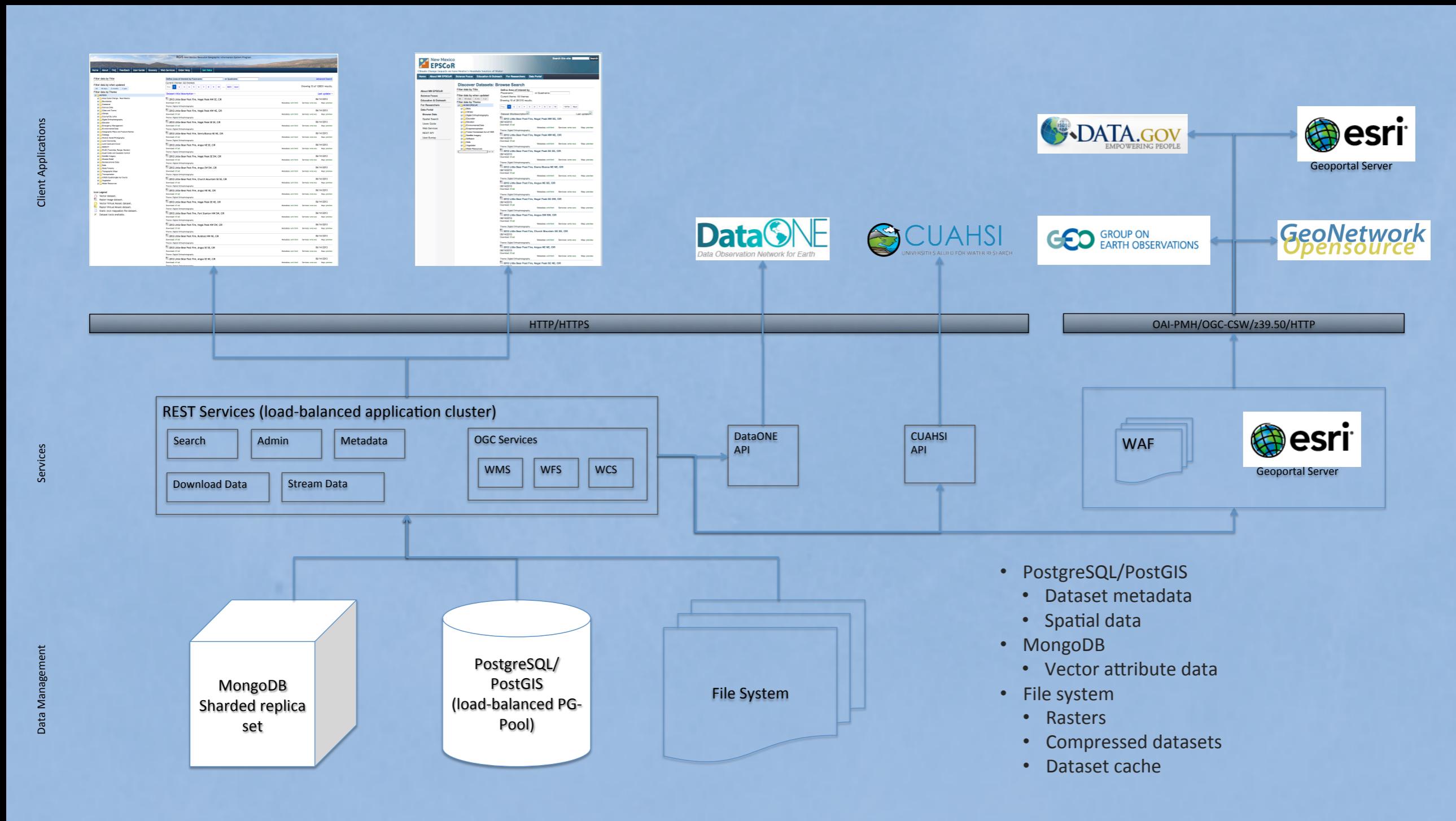
EDAC's Data Challenge

Our Solution

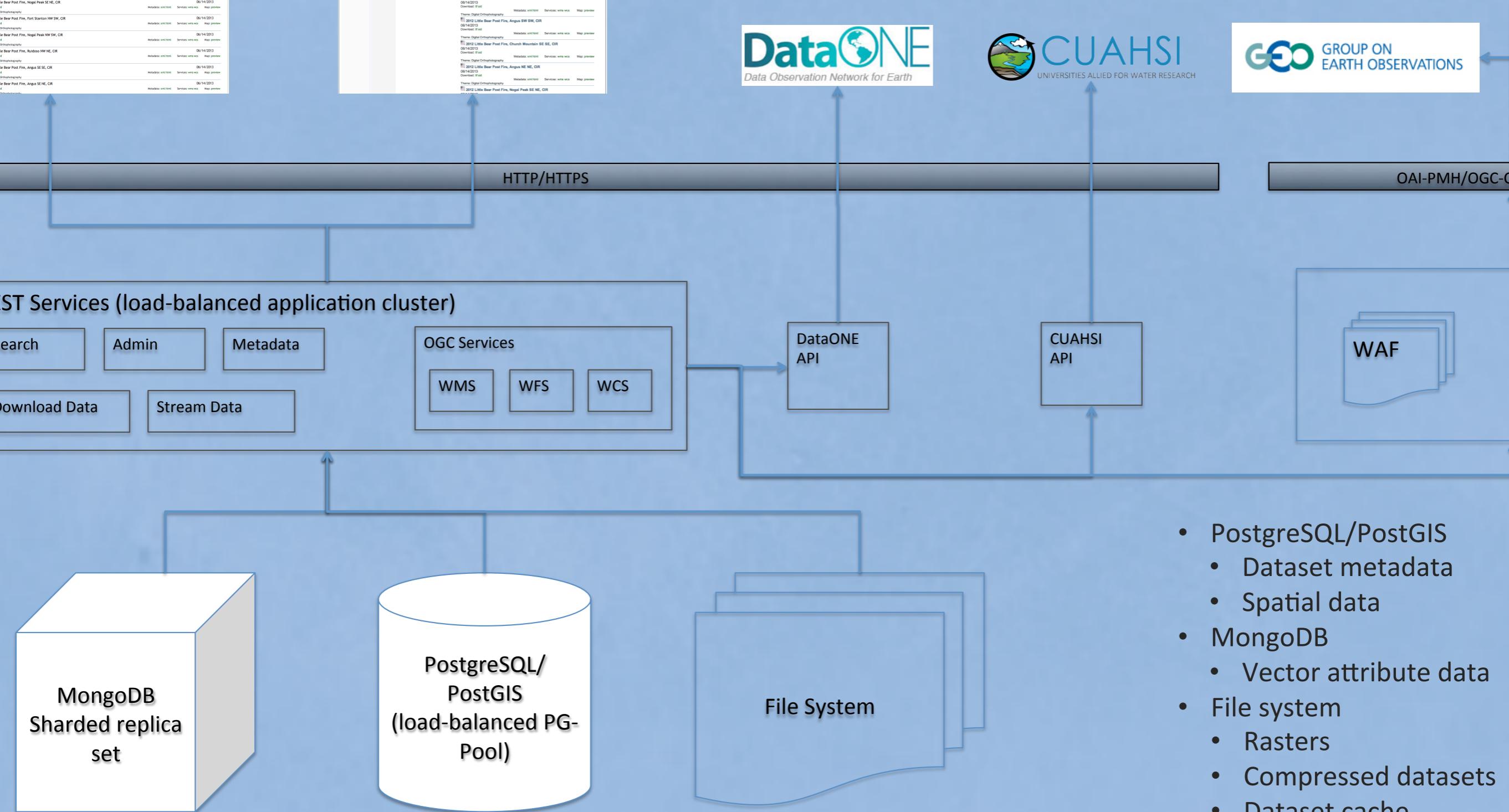


EDAC's Data Challenge

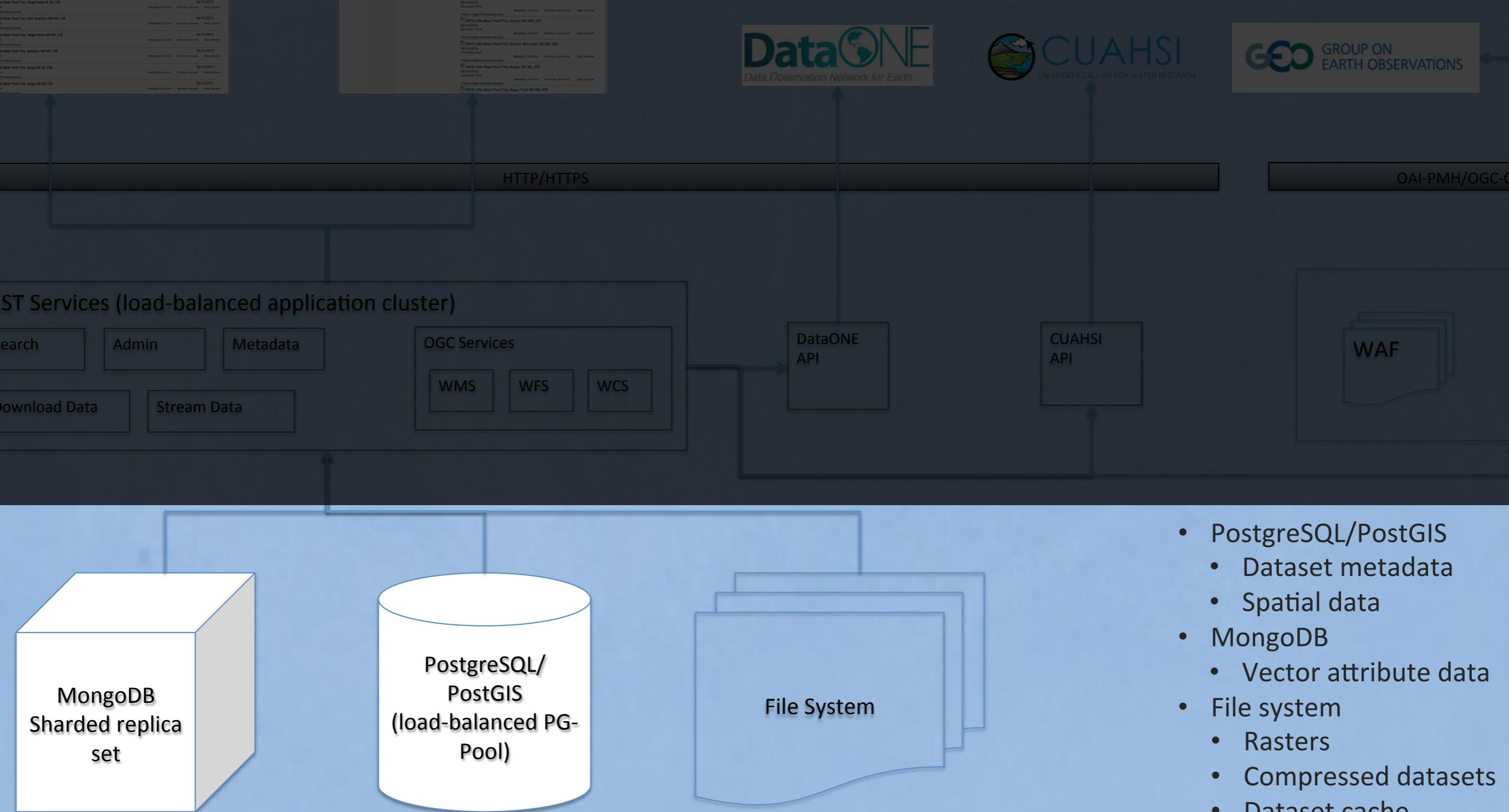
Our Solution



- PostgreSQL/PostGIS
 - Dataset metadata
 - Spatial data
- MongoDB
 - Vector attribute data
- File system
 - Rasters
 - Compressed datasets
 - Dataset cache



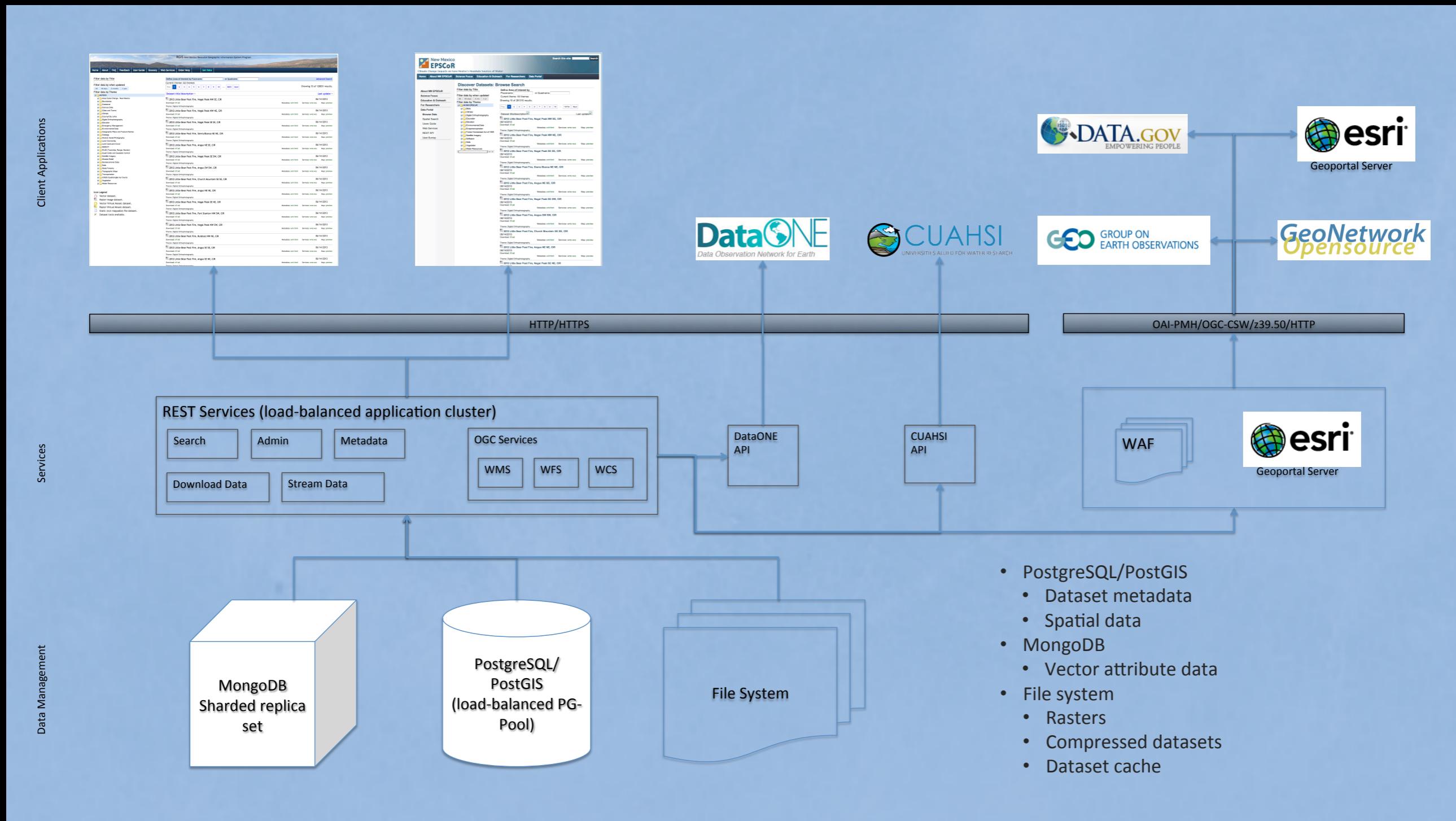
- PostgreSQL/PostGIS
 - Dataset metadata
 - Spatial data
- MongoDB
 - Vector attribute data
- File system
 - Rasters
 - Compressed datasets
 - Dataset cache



- PostgreSQL/PostGIS
 - Dataset metadata
 - Spatial data
- MongoDB
 - Vector attribute data
 - File system
 - Rasters
 - Compressed datasets
 - Dataset cache

EDAC's Data Challenge

Our Solution



2012 Little Bear Post Fire, Angus SE SE, CIR
Download: tf id
Theme: Digital Orthophotography
2012 Little Bear Post Fire, Angus SE NE, CIR
Download: tf id
Theme: Digital Orthophotography

06/14/2013
Metadata: xml html Services: wms wcs Map: preview
2012 Little Bear Post Fire, Nogal Peak SE NE, CIR
Download: tf id
Theme: Digital Orthophotography

2012 Little Bear Post Fire, Angus NE NE, CIR
06/14/2013
Download: tf id
Theme: Digital Orthophotography
2012 Little Bear Post Fire, Nogal Peak SE NE, CIR
Download: tf id
Theme: Digital Orthophotography

REST Services (load-balanced application cluster)

Search

Admin

Metadata

Download Data

Stream Data

OGC Services

WMS

WFS

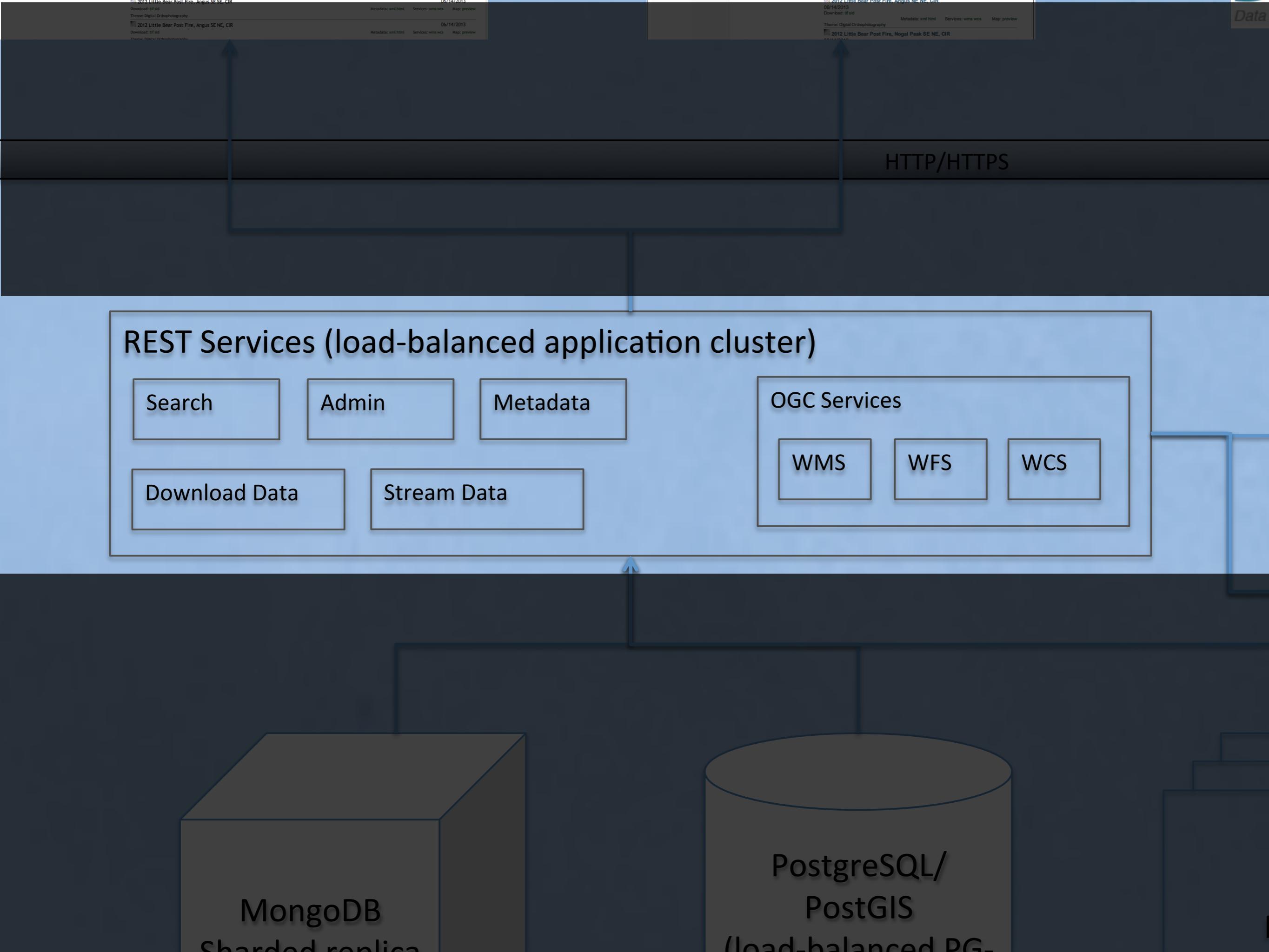
WCS

MongoDB
Sharded replica

PostgreSQL/
PostGIS
(load-balanced PG-

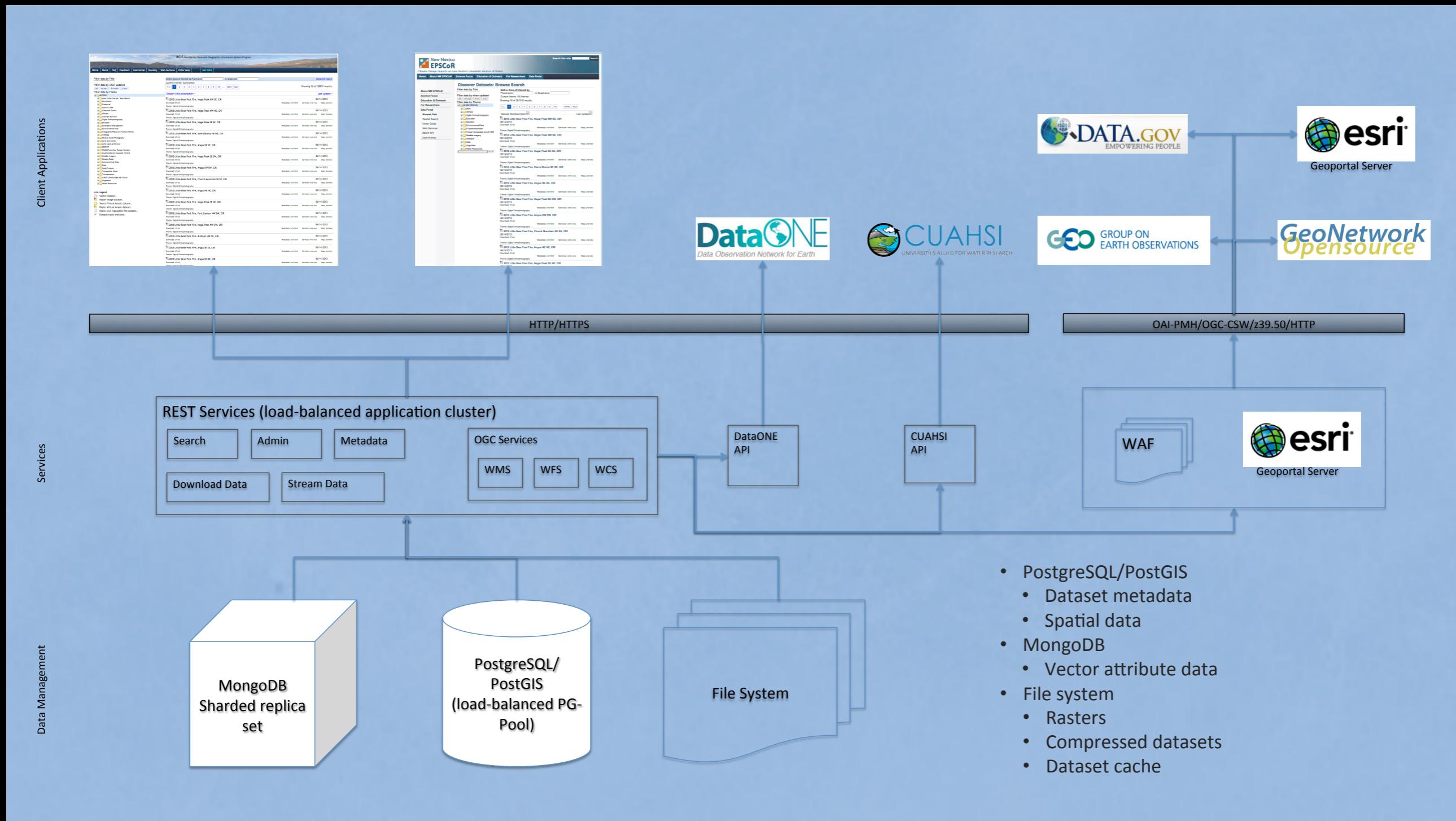
P

HTTP/HTTPS



EDAC's Data Challenge

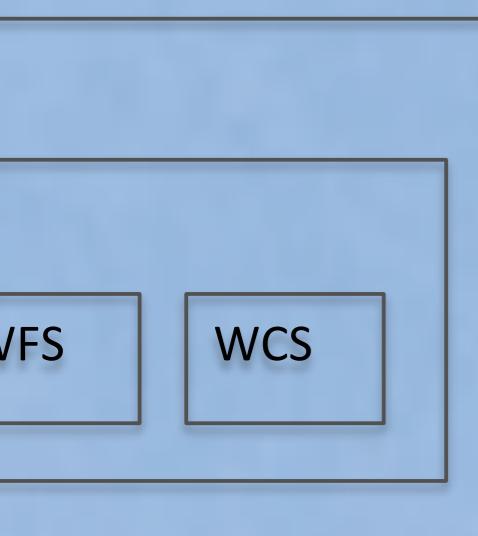
Our Solution



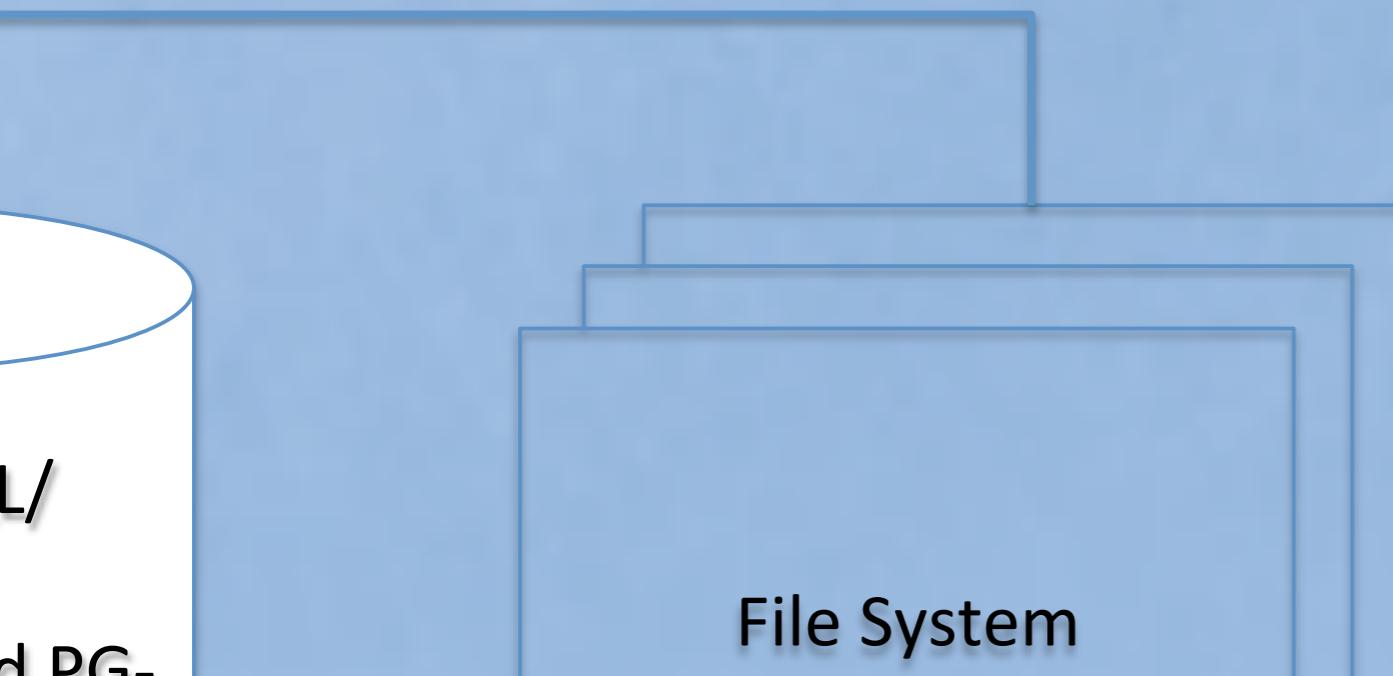
- PostgreSQL/PostGIS
 - Dataset metadata
 - Spatial data
- MongoDB
 - Vector attribute data
- File system
 - Rasters
 - Compressed datasets
 - Dataset cache

HTTP/HTTPS

O

DataONE
APICUAHSI
API

WAF

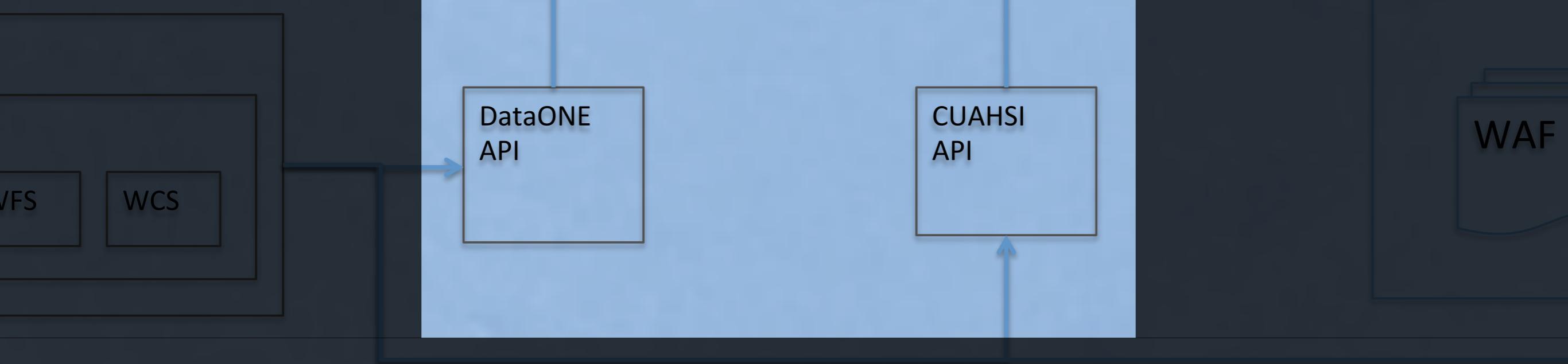


File System

- PostgreSQL/PostGIS
 - Dataset metadata
 - Spatial data
- MongoDB
 - Vector attributes
- File system

FTP/HTTPS

O

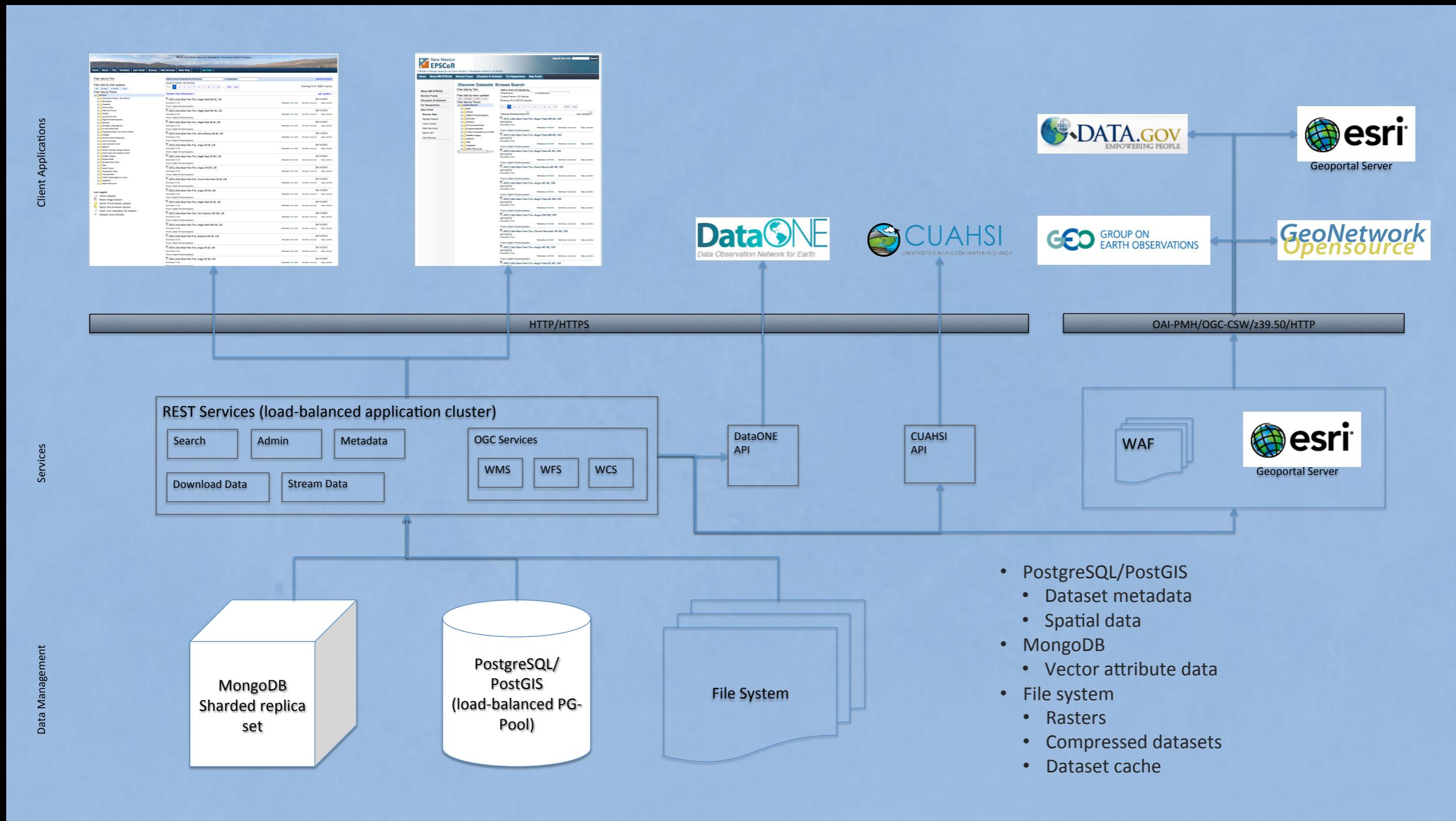
L/
d PG-

File System

- PostgreSQL/PostGIS
 - Dataset metadata
 - Spatial data
- MongoDB
 - Vector attributes
- File system

EDAC's Data Challenge

Our Solution



- PostgreSQL/PostGIS
 - Dataset metadata
 - Spatial data
- MongoDB
 - Vector attribute data
- File system
 - Rasters
 - Compressed datasets
 - Dataset cache



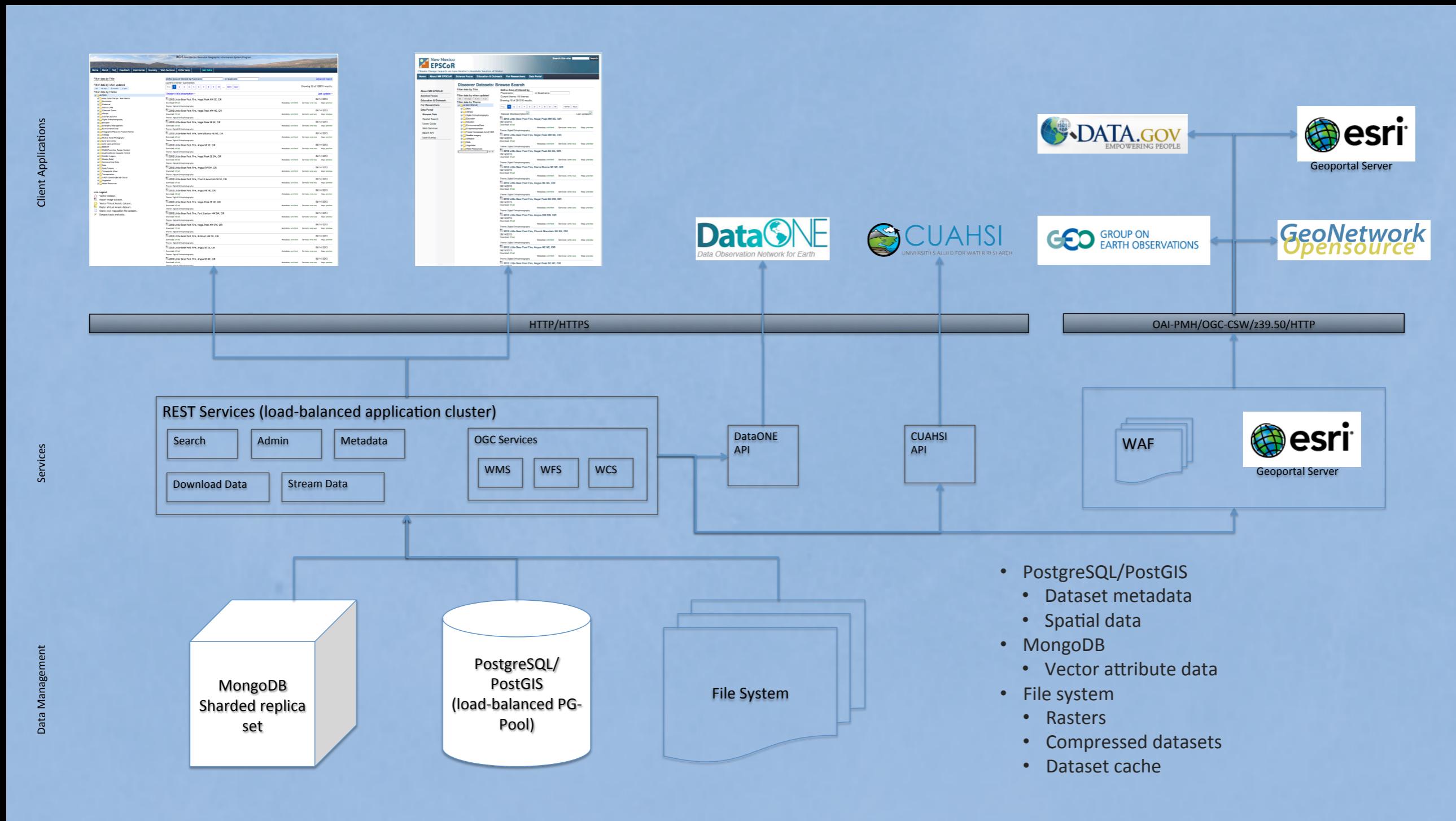
- PostgreSQL/PostGIS
 - Dataset metadata
 - Spatial data
- MongoDB
 - Vector attribute data
- File system



- PostgreSQL/PostGIS
 - Dataset metadata
 - Spatial data
- MongoDB
 - Vector attribute data
- File system

EDAC's Data Challenge

Our Solution



Client Applications

RGIS New Mexico Resource Geographic Information System Program

Home | About | FAQ | Feedback | User Guide | Glossary | Web Services | Order Help | Get Data

Filter data by Title
Define Area of Interest by Placename or Quadname Advanced Search

Filter data by when updated
All 90 days 6 months 1 year

Filter data by Theme
All RGIS

- Area Code Change - New Mexico
- Boundaries
- Cadastral
- Census Data
- Cities and Towns
- Climate
- County/City Links
- Digital Orthophotography
- Elevation
- Emergency Management
- Environmental Data
- Geographic Place and Feature Names
- Geology
- Historic Aerial Photography
- Land Ownership
- Land Use/Land Cover
- NMDOT
- PLSS (Township, Range, Section)
- Quad Grids and Geodetic Control
- Satellite Imagery
- Shaded Relief
- Socioeconomic Data
- Soils
- State Forestry
- Topographic Maps
- Transportation
- USGS Quadrangles by County
- Vegetation
- Water Resources

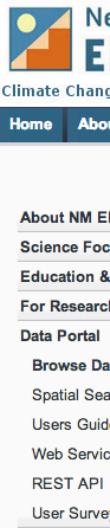
Icon Legend

- Vector dataset.
- Raster image dataset.
- Vector Virtual Mosaic dataset.
- Raster Virtual Mosaic dataset.
- Static (non mappable) file dataset.
- Dataset tools available.

Current theme: All themes
Prev 1 2 3 4 5 6 7 8 9 10 ... 8003 Next

Showing 15 of 120031 results.

Dataset title/description	Last update
2012 Little Bear Post Fire, Nogal Peak NW SE, CIR	06/14/2013
Download: tif sid Theme: Digital Orthophotography	Metadata: xml html Services: wms wcs Map: preview
2012 Little Bear Post Fire, Nogal Peak NW NE, CIR	06/14/2013
Download: tif sid Theme: Digital Orthophotography	Metadata: xml html Services: wms wcs Map: preview
2012 Little Bear Post Fire, Nogal Peak SE SE, CIR	06/14/2013
Download: tif sid Theme: Digital Orthophotography	Metadata: xml html Services: wms wcs Map: preview
2012 Little Bear Post Fire, Sierra Blanca NE NE, CIR	06/14/2013
Download: tif sid Theme: Digital Orthophotography	Metadata: xml html Services: wms wcs Map: preview
2012 Little Bear Post Fire, Angus NE SE, CIR	06/14/2013
Download: tif sid Theme: Digital Orthophotography	Metadata: xml html Services: wms wcs Map: preview
2012 Little Bear Post Fire, Nogal Peak SE SW, CIR	06/14/2013
Download: tif sid Theme: Digital Orthophotography	Metadata: xml html Services: wms wcs Map: preview
2012 Little Bear Post Fire, Angus SW SW, CIR	06/14/2013
Download: tif sid Theme: Digital Orthophotography	Metadata: xml html Services: wms wcs Map: preview
2012 Little Bear Post Fire, Church Mountain SE SE, CIR	06/14/2013
Download: tif sid Theme: Digital Orthophotography	Metadata: xml html Services: wms wcs Map: preview
2012 Little Bear Post Fire, Angus NE NE, CIR	06/14/2013
Download: tif sid Theme: Digital Orthophotography	Metadata: xml html Services: wms wcs Map: preview
2012 Little Bear Post Fire, Nogal Peak SE NE, CIR	06/14/2013
Download: tif sid Theme: Digital Orthophotography	Metadata: xml html Services: wms wcs Map: preview
2012 Little Bear Post Fire, Fort Stanton NW SW, CIR	06/14/2013
Download: tif sid Theme: Digital Orthophotography	Metadata: xml html Services: wms wcs Map: preview
2012 Little Bear Post Fire, Nogal Peak NW SW, CIR	06/14/2013
Download: tif sid Theme: Digital Orthophotography	Metadata: xml html Services: wms wcs Map: preview
2012 Little Bear Post Fire, Ruidoso NW NE, CIR	06/14/2013
Download: tif sid Theme: Digital Orthophotography	Metadata: xml html Services: wms wcs Map: preview
2012 Little Bear Post Fire, Angus SE SE, CIR	06/14/2013
Download: tif sid Theme: Digital Orthophotography	Metadata: xml html Services: wms wcs Map: preview
2012 Little Bear Post Fire, Angus SE NE, CIR	06/14/2013
Download: tif sid Theme: Digital Orthophotography	Metadata: xml html Services: wms wcs Map: preview



Showing 15 of 120031 results.		
Last update	Services: wms wcs	Map: preview
06/14/2013	xml html	Services: wms wcs Map: preview
06/14/2013	xml html	Services: wms wcs Map: preview
06/14/2013	xml html	Services: wms wcs Map: preview
06/14/2013	xml html	Services: wms wcs Map: preview
06/14/2013	xml html	Services: wms wcs Map: preview
06/14/2013	xml html	Services: wms wcs Map: preview
06/14/2013	xml html	Services: wms wcs Map: preview
06/14/2013	xml html	Services: wms wcs Map: preview
06/14/2013	xml html	Services: wms wcs Map: preview
06/14/2013	xml html	Services: wms wcs Map: preview
06/14/2013	xml html	Services: wms wcs Map: preview
06/14/2013	xml html	Services: wms wcs Map: preview
06/14/2013	xml html	Services: wms wcs Map: preview
06/14/2013	xml html	Services: wms wcs Map: preview

New Mexico EPSCoR

Climate Change Impacts on New Mexico's Mountain Sources of Water

Search this site:

Home | About NM EPSCoR | Science Focus | Education & Outreach | For Researchers | Data Portal

Discover Datasets: Browse Search

Filter data by Title
Filter data by when updated
All 90 days 6 mo 1 yr

Filter data by Theme
All NM EPSCoR

- Biota
- Climate
- Digital Orthophraphy
- Education
- Elevation
- Environmental Data
- Evapotranspiration
- Forest Homestead Act of 1906
- Satellite Imagery
- Software
- Soils
- Vegetation
- Water Resources

Showing 15 of 281310 results.

Dataset title/description Last update

 2012 Little Bear Post Fire, Nogal Peak NW SE, CIR
06/14/2013
Download: [tif sid](#) Metadata: [xml html](#) Services: [wms wcs](#) Map: [preview](#)
Theme: Digital Orthophraphy

 2012 Little Bear Post Fire, Nogal Peak NW NE, CIR
06/14/2013
Download: [tif sid](#) Metadata: [xml html](#) Services: [wms wcs](#) Map: [preview](#)
Theme: Digital Orthophraphy

 2012 Little Bear Post Fire, Nogal Peak SE SE, CIR
06/14/2013
Download: [tif sid](#) Metadata: [xml html](#) Services: [wms wcs](#) Map: [preview](#)
Theme: Digital Orthophraphy

 2012 Little Bear Post Fire, Sierra Blanca NE NE, CIR
06/14/2013
Download: [tif sid](#) Metadata: [xml html](#) Services: [wms wcs](#) Map: [preview](#)
Theme: Digital Orthophraphy

 2012 Little Bear Post Fire, Angus NE SE, CIR
06/14/2013
Download: [tif sid](#) Metadata: [xml html](#) Services: [wms wcs](#) Map: [preview](#)
Theme: Digital Orthophraphy

 2012 Little Bear Post Fire, Nogal Peak SE SW, CIR
06/14/2013
Download: [tif sid](#) Metadata: [xml html](#) Services: [wms wcs](#) Map: [preview](#)
Theme: Digital Orthophraphy

 2012 Little Bear Post Fire, Angus SW SW, CIR
06/14/2013
Download: [tif sid](#) Metadata: [xml html](#) Services: [wms wcs](#) Map: [preview](#)
Theme: Digital Orthophraphy

 2012 Little Bear Post Fire, Church Mountain SE SE, CIR
06/14/2013
Download: [tif sid](#) Metadata: [xml html](#) Services: [wms wcs](#) Map: [preview](#)
Theme: Digital Orthophraphy

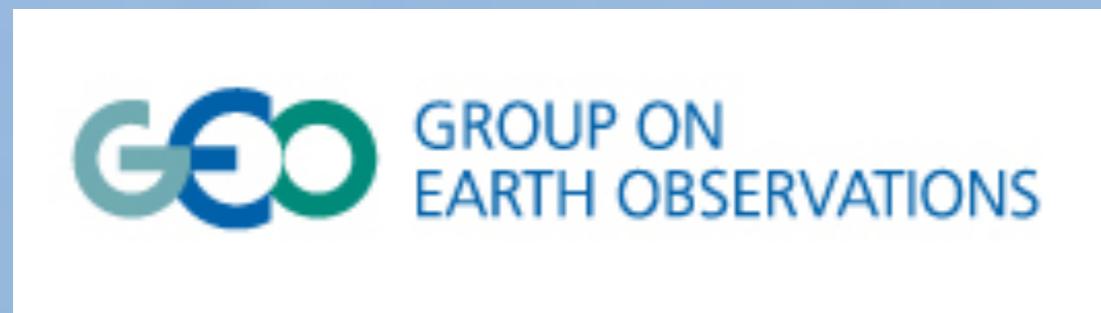
 2012 Little Bear Post Fire, Angus NE NE, CIR
06/14/2013
Download: [tif sid](#) Metadata: [xml html](#) Services: [wms wcs](#) Map: [preview](#)
Theme: Digital Orthophraphy

 2012 Little Bear Post Fire, Nogal Peak SE NE, CIR
06/14/2013
Download: [tif sid](#) Metadata: [xml html](#) Services: [wms wcs](#) Map: [preview](#)
Theme: Digital Orthophraphy

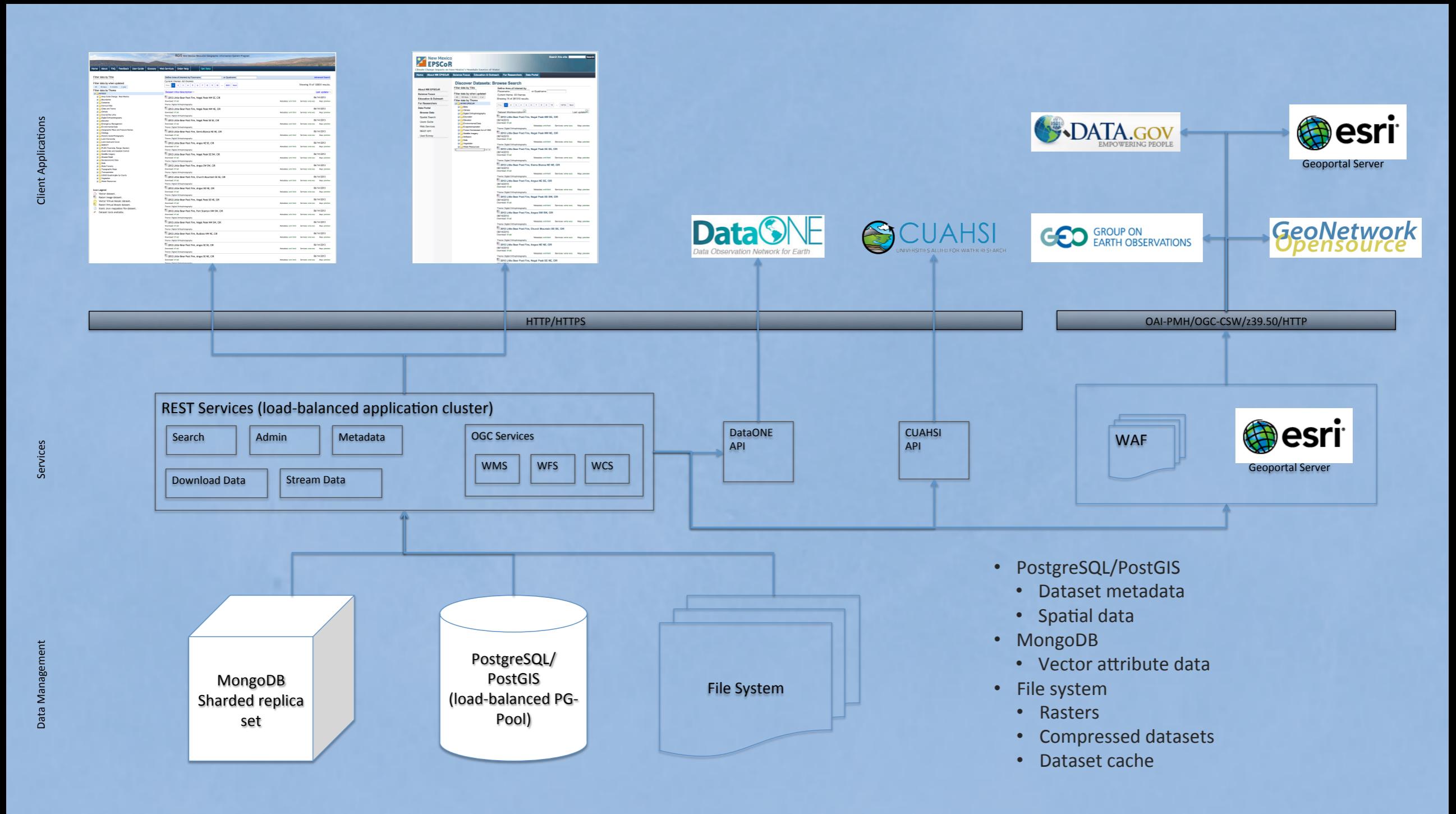
HTTP/HTTPS

D
Data C





Cloud Computing & Big Data?



Gstore/NM RGIS Demonstration

<http://rgis.unm.edu>
<http://gstore.unm.edu>

Conclusions



THE NETWORK WENT
DOWN AND I LOST
MY WORK.

THE SERVER
CRASHED.

FROM NOW ON, I
WANT ADVANCED
NOTICE OF ANY
UNPLANNED OUTAGES.

AND I
NEED IT
YESTERDAY.

I USED TO
THINK THAT
WAS JUST A
FIGURE OF
SPEECH.

S. Adams www.unitedmedia.com

1/28/97 © 1997 United Feature Syndicate, Inc.

Questions?



Acknowledgements

- The products and platforms upon which this work is based have been funded by
 - NSF EPSCoR Program (Track 1 [Awards: 0447691, 0814449, 1301346] and Track 2 awards [0918635, 1329470])
 - New Mexico Resource Geographic Information System
 - NASA ACCESS Program
- Unless otherwise noted, all photos are from shutterstock.com