

The National DATA SERVICE

- Towards a world where it is easier to *publish, link, search, and reuse* data of all kinds
 - Advancing discovery by enabling open sharing of data
 - Increase collaboration within/across fields
- National effort to bring together R&D efforts surrounding data
 - *Interoperability* and extensibility of data cyberinfrastructure efforts
 - In collaboration with the RDA implementing *standards* and protocols



RESEARCH DATA ALLIANCE

EASILY ...

- preserve data for longer periods of time
- find data
- share data collections over the web
- set data access permissions
- publish data, cite data, get credit for your data
- efficiently and reliably move large amounts of data and/or large numbers of files
- run custom tools/software on data, near the data
- run compute intensive analysis on data
- run analysis on large collections of data
- run analysis across any available resources
- utilize an extensible suite of reusable data analysis/manipulation tools
- preserve, share, and find software/tools associated with the data
- publish software, cite software, get credit for your software
- create and share workflows over data
- **ask novel questions spanning all available data**

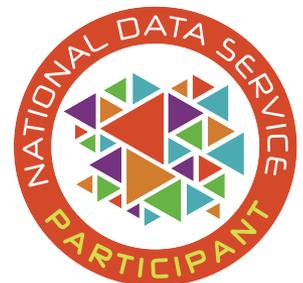
Enabling Discovery with Open Data

- Scientists have pinned down a molecular process in the brain that helps to trigger schizophrenia
- Analyzed data from about 29,000 schizophrenia cases, 36,000 controls and 700 post mortem brains. The information was drawn from dozens of studies performed in 22 countries.
- The authors stressed that their findings, which combine basic science with large-scale analysis of genetic studies, depended on an unusual level of cooperation among experts in genetics, molecular biology, developmental neurobiology and immunology.
- "This could not have been done five years ago".

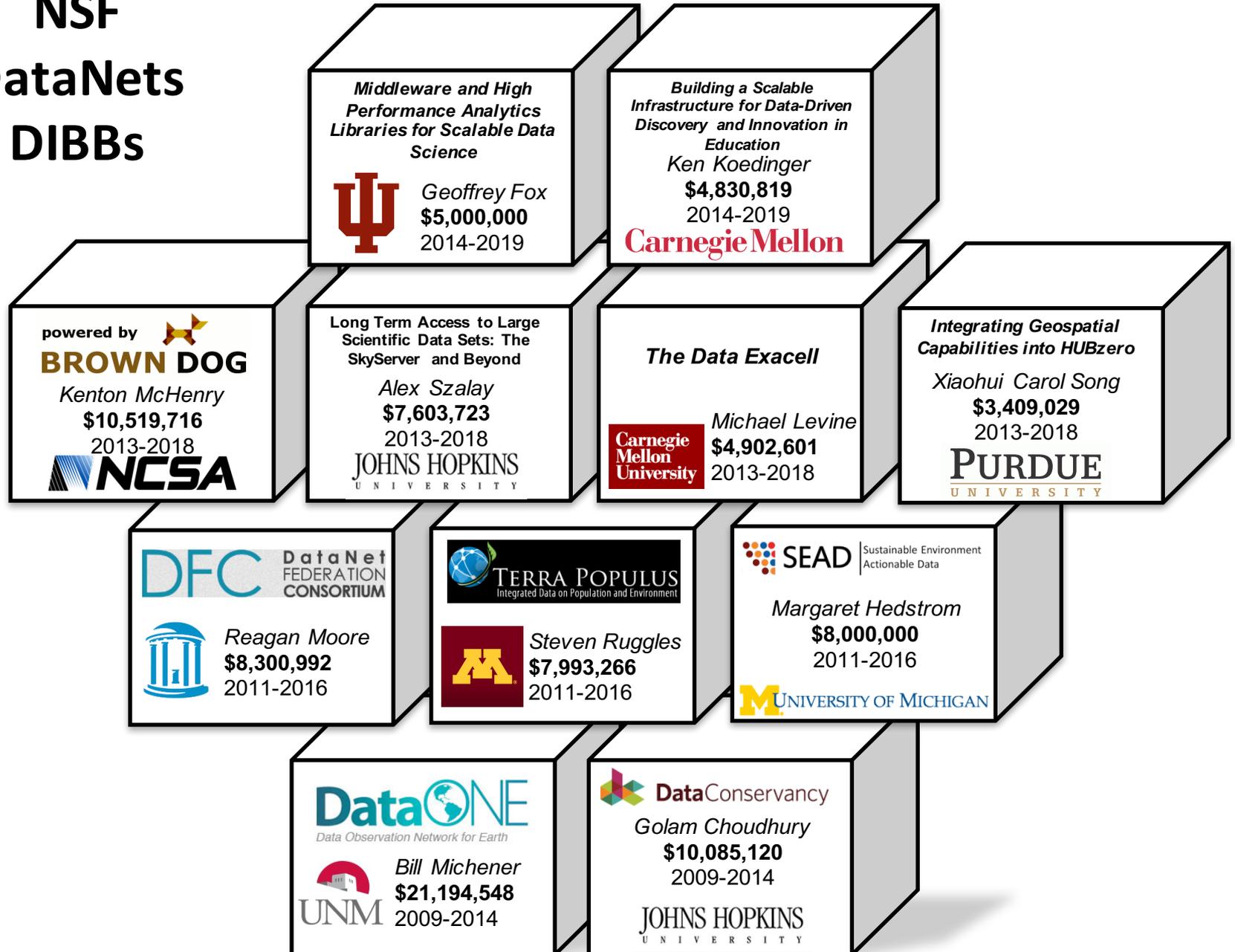
<https://www.washingtonpost.com/news/speaking-of-science/wp/2016/01/27/scientists-open-the-black-box-of-schizophrenia-with-dramatic-genetic-finding/>

The National Data Service Consortium

- Collaboration of many kinds of institutions
 - Computational and data service centers
 - ANL, NCSA, PSC, SDSC, TACC
 - Universities and library repositories
 - CU Boulder, Harvard, Indiana, Purdue, UIUC
 - Community efforts
 - LSST, LIGO, ICPSR, ...
 - Cyberinfrastructure efforts
 - DataONE, iPlant, IRODs, Globus, SciServer, SEAD, ...
 - Publishers
 - Nature, Science, APS, IEEE, PLOS, Elsevier, ...
- Consortium to guide the building/governance of services
 - Bi-annual workshops
 - NDS Consortium Steering Committee
 - Coordinate separately funded efforts to build NDS components
 - Ensure interoperability, integrate existing tools and resources
- Technical Advisory Committee
 - technical guidance to development of integrative services



NSF DataNets DIBBs



Midwest Big Data Hub

Potential Tools, Services, and Resources

This page collects information about potential tools, services and resources in the MidWest region (Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin) with the potential to support MBDH spokes and other activities. This page is being updated as information is collected. To add or amend entries please write to admin@midwestbigdatahub.org.

These resources are in addition to the [National Data Service \(NDS\)](#) capabilities which are available for spokes and supported by the MBDH as discussed on the hub web pages. The resources listed here were identified as assets available within our region. Spoke PIs should contact the organizations and individuals listed below as point of contacts to discuss availability of services.

[Networks](#)
[HPC Centers and Compute Resources](#)
[Cloud Resources](#)
[Data Curation and Publishing Tools](#)
[Storage and Repositories](#)
[Data Facilities and Infrastructure](#)
[Services](#)
[Software Services](#)
[Specialists](#)
[Educational Resources](#)
[Tech Transfer and/or Incubators](#)
[Experimental facilities that generate data](#)

Networks

- [BOREAS](#), Broadband Optical Research, Education and Sciences Network (Iowa, Minnesota, Wisconsin)
 - Collaboration of Iowa State University, University of Iowa, University of Minnesota, and University of Wisconsin-Madison.
- [Citizen Science Grid](#) (North Dakota), [Travis Desell](#)
 - Hosted by UND's Computational Research Center and Information Technology Systems and Services, dedicated to supporting a wide range of research and educational projects using volunteer computing and citizen science.
- [Great Plains Network](#) (Iowa, Kansas, Minnesota, Missouri, North Dakota, South Dakota, Wisconsin)
 - PoPs in Kansas City, Kansas, and Tulsa, Oklahoma; connects 9 state networks to the Internet2 and ESnet backbones:

**89 regional
resources**

Break Down Science Drivers

- Data Collections
 - Un-digitized
 - Offline
 - Restricted
 - Unstructured
 - Un-curated
 - Simulation outputs
 - Representations
 - Diverse data
 - Cross disciplinary collections
 - Large datasets
 - Many small datasets
- Data Handling Needs
 - Data hosting
 - Access control
 - Data collaboration
 - Data curation tools
 - Data analysis tools
 - Data visualizations
 - Long term archiving
 - Support for community tools
 - High performance computation
 - Cloud computation

Break Down of Technology Components

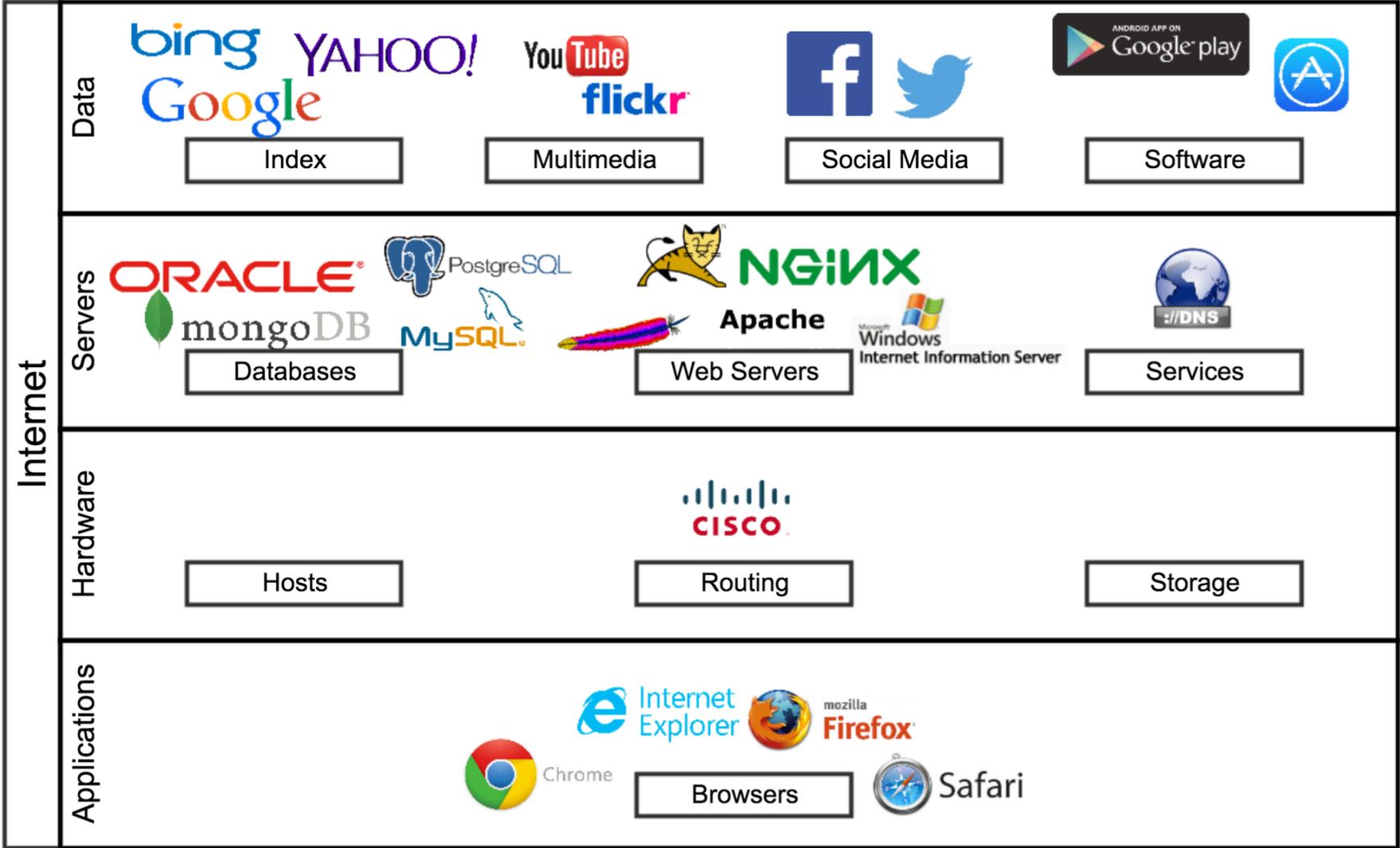
- Infrastructure & Resources
 - **Data**
 - Networking
 - Storage
 - Security
 - High performance computing
 - Cloud computing resources
- Data Service Capabilities
 - Authentication
 - Identity management
 - Access control
 - Transfer
 - Large transfers
 - Robust transfer
 - Storage
 - Storage abstraction
 - Sharing/collaboration
 - Archiving
 - Replication
 - Data publishing
 - Code publishing
 - Curation
 - Ingestion
 - Description
 - Provenance
 - Transformation
 - Integration
 - Analysis
 - Data analysis
 - Event notification
 - Exploration
 - Federated search
 - API
 - User interface

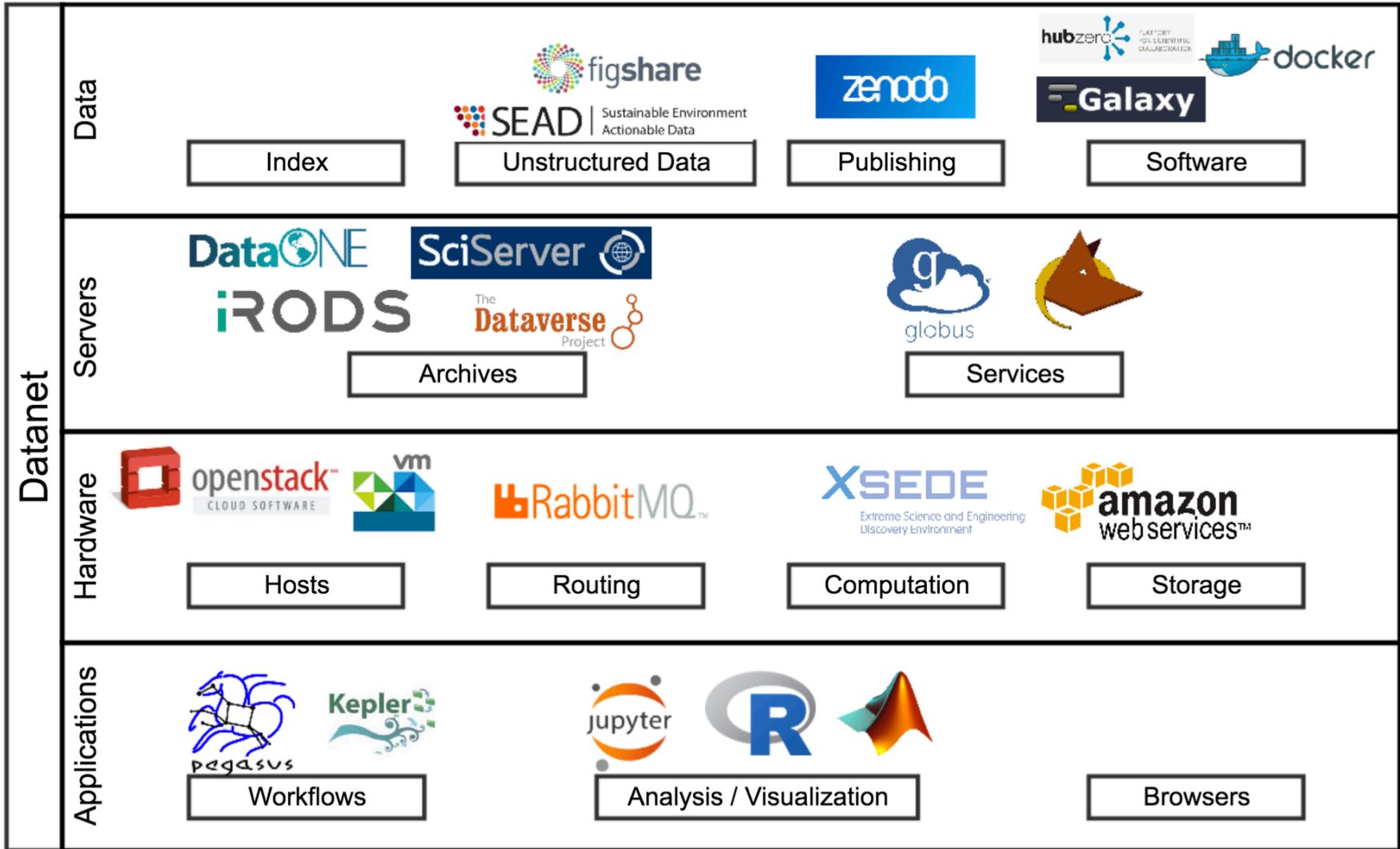
U.S. NDS Components

Data Service Capabilities												
			Curation					Analysis		Exploration		
	Data Publishing	Code Publishing	Data Ingestion/ Curation	Data Description	Provenance	Data Transformation	Data Integration	Data Analysis	Event Notification	Federated Search	Data Exploration API	Data Exploration User Interface
BD2K							○					
Blue Waters												
Brown Dog		○			○	○		○			○	
CARMA												
CyberGIS							○	○				
CyVerse	○		○									
DataONE	○		○		○						○	○
Dataverse												
Globus	○		○									
HUBzero		○						○				
iRODS												
Materials Data Facility			○	○								
NIST-CORE							○	○				

...







Tools & Services

- Contained broadly reusable components addressing some specific data need
 - Data Transfer Service
 - Identity Service
 - Data Storage Service
 - HPC/Cloud Computation Service
 - Data Sharing/Publishing Services
 - Data Archiving Service
 - Data Transformation Service
 - Code/Tool Publishing Service
- ***Interoperability with other such components***

NDSLabs Demo Sign Out ↗

Stacks Volumes

- Clowder +
- Dataverse +
- ElasticSearch +
- iRODS Clowdbrowser UI +
- iRODS iCAT +
- Kibana (ELK) +
- MongoDB +
- PostgreSQL +
- RabbitMQ +
- Rserve +

✔
Dataverse-01
▼

Status	Volumes	Name	Config	Logs
✔ Ready		Dataverse ↗		
✔ Ready		PostgreSQL		
✔ Ready		Solr		
✔ Ready		Rserve		
✔ Ready		Dataverse iRODS iCAT		

■ Stop Stack

✔
Irods clowdb-01
▼

Status	Volumes	Name	Config	Logs
--------	---------	------	--------	------

<https://www.youtube.com/channel/UCWuPo7LDCzsqF3RaKzIQmHw>





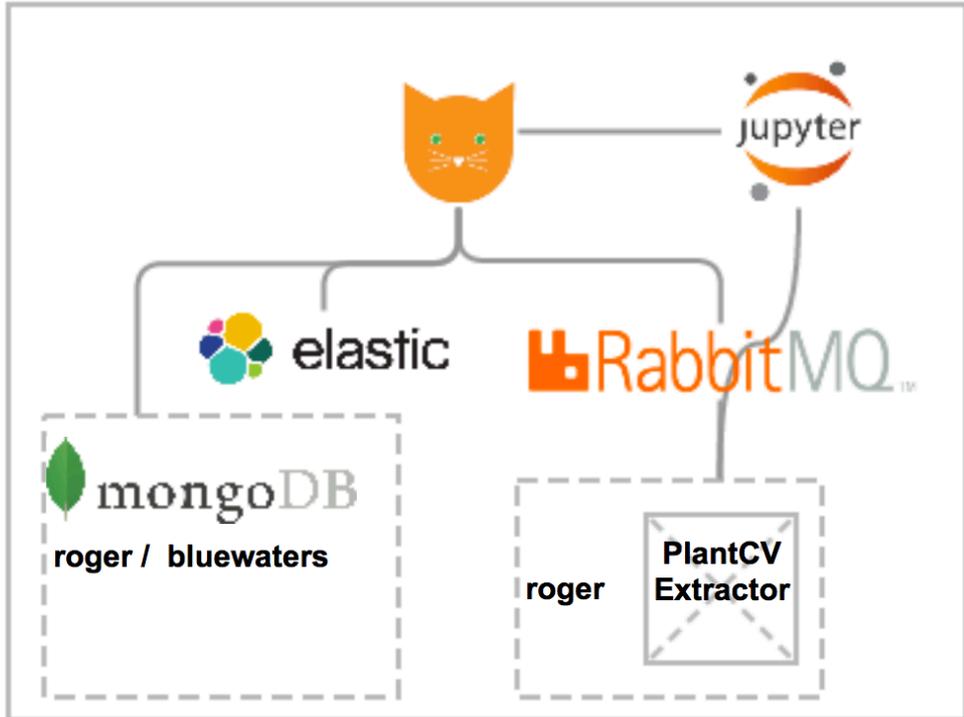
Search NDS services

- apache
- clowder
- elastic
- gitlab
- globus
- irods
- jupyter

Search NDS resources

- amazon
- bluewaters
- chameleon
- comet
- jetstream
- roger

TERRA Deployment



Commit