Towards a U.S. National Data Service

Inaugural Report

January 1st to June 30th 2016

John Towns, Interim Director
Christine Kirkpatrick, Chair, Technical Advisory Committee
Kenton McHenry, Technical Coordinator
Kandace Turner, Project Manager
http://nationaldataservice.org

The National Data Service Consortium[1] (NDSC) aims to pilot and bring together activities necessary towards the establishment of a U.S. National Data Service (NDS), analogous to those currently underway in Australia[2] and the European Union[3,4]. At a high level, this entails not only the storage and computation resources required by contemporary scientific research involving digital data, but also the services on top of those resources for metadata, curation, indexing, storage abstraction, replication, data transfer, authentication, access control, transformation, analysis and reproducible workflows. An established NDS allows for scientific data accessibility - to be published and linked with the paper publication, allowing others to search for and reuse the data towards further novel discoveries[5]. This report outlines the development, organizational, and community engagement activities undertaken by the NDSC towards this goal during its inception: highlighting January 1 - June 30th 2016.

[1] http://www.nationaldataservice.org/

[2] http://www.ands.org.au/

[3] https://www.eudat.eu/

[4] https://ec.europa.eu/digital-single-market/en/%20european-cloud-initiative

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

Table of Contents

Executive Summary

The first biannual National Data Service report comes two years into NDS's inception. After a year of consortium building and stakeholder consultation, the NDS spent its next year extending its NDS Labs offering and refining it through pilot feedback. The Steering Committee (SC), the primary body representing Consortium institutions, produced a comprehensive landscape analysis, and advised on possible sustainability directions and future partnerships. The SC was the driving organizing force behind the NDS workshops in San Diego and Chapel Hill. The Technical Advisory Committee (TAC) provided a technical perspective on the need for NDS, pilot outreach, drafted an interoperability task force proposal, and gave feedback on communication platforms and support mechanisms. The Executive Committee provided vision and guidance to the NDS development and project team, cultivated funding opportunities, and recruited an Executive Director. The next period for NDS will build on this foundation by using community requirements to drive NDS Labs development, deploying researcher-facing tools via NDS Share, and being a critical player in an emerging national data strategy.

A comprehensive survey of the data management landscape shows that the National Data Service Consortium (NDSC) is active in most areas. NDS continued to participate in RDA[6]-sponsored events and working groups. At the NDS5 workshop, NSF DataNets and DIBBs projects attended and discussed creating deeper interoperability between projects, with the potential of NDS as a connecting fabric. The NSF DIBBs Whole Tale (WT) project was awarded, born from the NDSC and its shared vision of the need to collocate publications with research data and associated analysis (code). The NDS team strengthened its ties with and understanding of EarthCube, with initial efforts to use Earthcube's CI components and to pilot Earthcube activities using NDS products. The NSF Big Data Innovation Hubs were established with NDS members doing outreach to the Hub communities and work began on a Hub-wide infrastructure strategy. Additional capabilities were added to NDS Labs, most notably an NDS Labs Workbench that allows rapid prototyping and deployment of data services in the Labs environment. Other notable NDS Labs achievements included updated services, a developers API, and a Kubernetes architecture for orchestration and deployment. NDS Share planning continued with refinement of its vision, a draft charter, and an agreement for those providing resources. A number of pilots were undertaken in diverse areas such as materials science, astrophysics, and a knowledge network for biomedical scientists.

[6] https://rd-alliance.org/

1 Data Management Landscape

Over the past few decades, there has been significant investment in the development of applications, services, and tools for the management of scientific data. While this is still an active area of exploration, these existing tools have addressed data management needs in terms of data sharing, curation of digital artifacts towards long term preservation and reuse, diverse data types, data movement, security, analysis, and publication at a level analogous to that of scientific papers. While gaps still exist with regards to scientific needs and current capabilities, there is now a large and diverse landscape of software supporting scientific data management.

Scientific teams looking to take advantage of these data management tools must navigate this landscape which includes tools with a spectrum of capabilities and varying degrees of accessibility. Many tools are developed in, and are somewhat isolated to, specific scientific domains. Further, while many of the capabilities that would make up an end-to-end scientific data management pipeline, from data creation to publication, are touched on by currently existing tools, such a pipeline is never fully realized due to a number of factors that limit their usability and reach. Factors such as needing to choose a subset of tools from those with overlapping capabilities as well as the need for custom development to integrate these components make many of these pipelines not only difficult to assemble and sustain but one-offs tailored to the needs of a specific project or community.

An analogy from the Research Data Alliance (RDA) compares the current situation faced by scientists with data management tools to that of early Internet users, before standard protocols such as TCP/IP and HTTP or markup languages such as HTML were widely adopted (Figure 1).

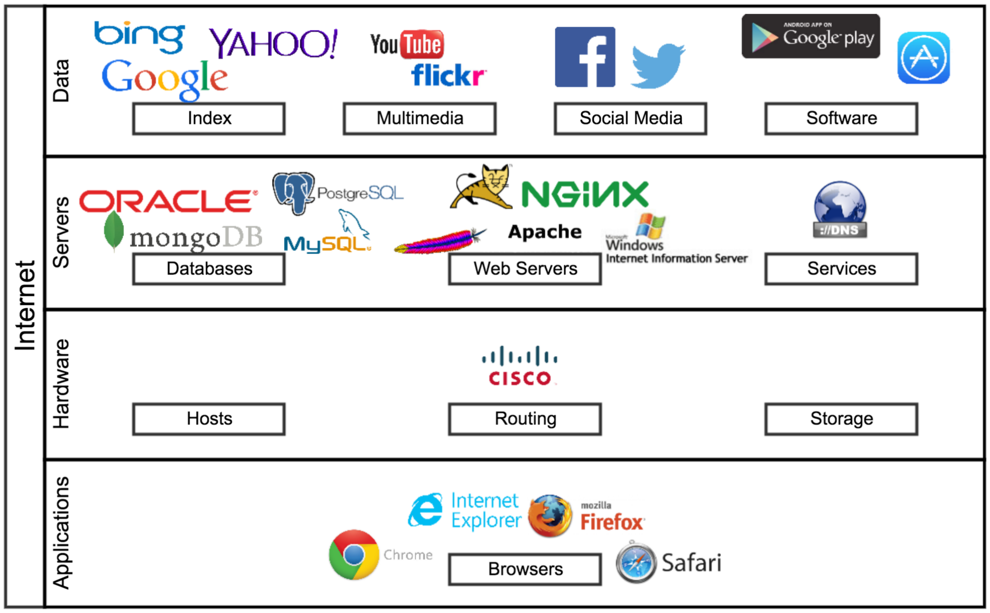


Figure 1: Array of Internet Technologies by Category

While Internet technologies vary widely, with many choices available for each component, this tends to not affect the user in terms of overall capabilities. A user can choose one of any number of browsers and still interoperate with other web technologies as they communicate via standard protocols and interfaces such as

HTTP and HTML. This is similar for administrators of web servers where any of a number of server software options can be chosen based on the administrators' familiarity with the software or desired performance/feature trade offs. Data management software is similar with archive software filling a role analogous to that of a web server and portals filling the role of browsers with a key distinction being the lack of standard protocols and interfaces (Figure 2).

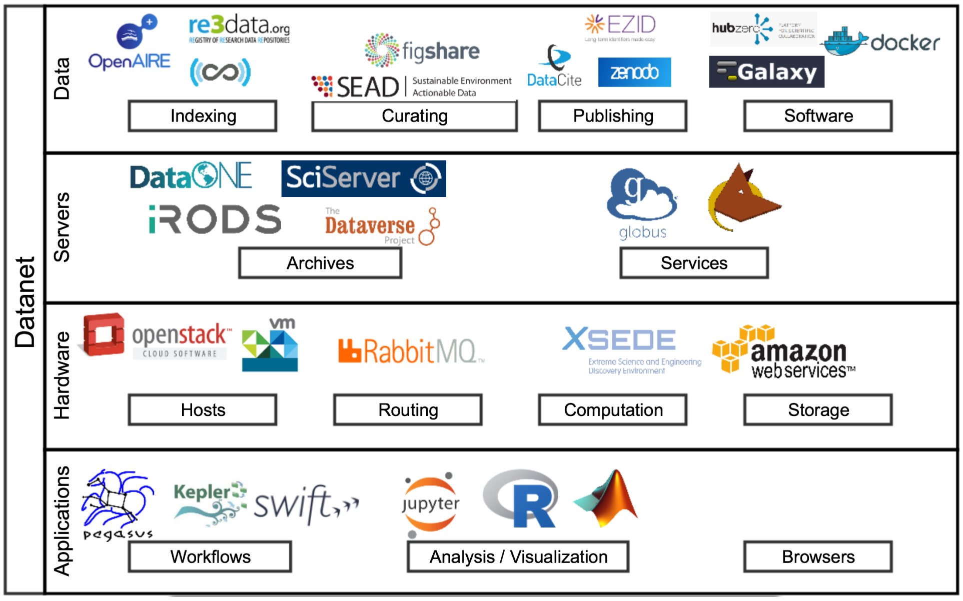


Figure 2: Array of Data Management Technologies by Category

The NDSC aims to foster the establishment of a broadly usable and sustainable data management pipeline and a National Data Service by directly addressing interoperability between developed data management components and pursuing the implementation of standardized interfaces between them[7].

1.1 Support of Pilot Activities

One mechanism being utilized by the NDSC towards fostering exposure and interoperability of data management components is in the support of pilot efforts that implement such connections. Initiated via a short proposal outlining the activity and its relevance towards a National Data Service[8], submissions are then reviewed by the NDS SC and TAC, made up of members of the scientific cyberinfrastructure community. If approved, the pilot effort is provided with development support as well as resources.

Numerous pilot efforts and collaborative development activities were undertaken over the past year (Section 3). Beyond supporting the individual pilots' research efforts, these activities facilitate a better understanding of the canonical components of a NDS and enumerate the options and gaps where they exist. The NDS iterated on a breakdown of some of these collaborative efforts[9] into a number of canonical service categories supported (e.g. identity management, data transfer, curation, transformation, analysis). This analysis assists in identifying gaps and areas with multiple offerings. This in turn informs future pilot efforts and convergence on standardized interfaces, interoperability mechanisms and enabling of a data management pipeline.

[7] http://www.nationaldataservice.org/docs/Charter_v2.pdf

[8] http://goo.gl/forms/uObA1cDJIUE02gqz2

[9] *https://nationaldataservice.atlassian.net/wiki/display/NDSC/NDS+Technology+Components*

1.2 Research Data Alliance

The NDS has worked closely with the RDA towards driving the establishment of policies, standards, and tools that can in turn be adopted and implemented within data management software components. Furthering this connection between policy, standards, and implementation, a National Data Service Interest Group (IG)[10] convened at the 7th RDA Plenary Meeting[11] in Tokyo where NDS activities across a number of nations described and compared their approaches and activities. Among the action items derived from this meeting was the sharing of the component breakdown matrix drafted by the U.S. NDS as a possible starting point for a version spanning international tools and services.[12] In addition to the NDS IG, the RDA has supported NDS pilot activities that bring these elements together[13].

1.3 NSF DataNets and DIBBs

The National Science Foundation (NSF) made a number of large investments in community building and research and development efforts that address scientific data management needs. Two programs in particular, DataNET[14] and Data Infrastructure Building Blocks (DIBBs)[15], led to the creation of a suite of complementary components that if joined together could make up a data management pipeline. The 5th NDS Workshop highlighted these activities with demonstrations of almost all DIBBS and DataNET tools as well as participation from the projects' representatives. Working closely with a number of these efforts, NDS is exploring how these components might be brought together[16].

Building on Essawy et al.[17], which described a scenario involving the interconnection of the DataNET efforts towards supporting a specific scientific use case, it is possible to conceive of an end-to-end data management pipeline built on the interconnection of these components through a handful of interfaces that would need to be standardized (Figure 3).

[10] https://rd-alliance.org/groups/national-data-services.html

[11] https://rd-alliance.org/plenaries/rda-seventh-plenary-meeting-tokyo-japan

[12] https://rd-alliance.org/group/national-data-services/wiki/national-data-service-efforts

[13] http://us.rd-alliance.org/news/rdaus-2016-call-adoption-projects

[14] https://en.wikipedia.org/wiki/Datanet

[15] https://en.wikipedia.org/wiki/Data_Infrastructure_Building_Blocks_(DIBBs)

[16] https://nationaldataservice.atlassian.net/wiki/pages/viewpage.action?pageId=4685968

[17] http://onlinelibrary.wiley.com/doi/10.1002/2015EA000139/pdf

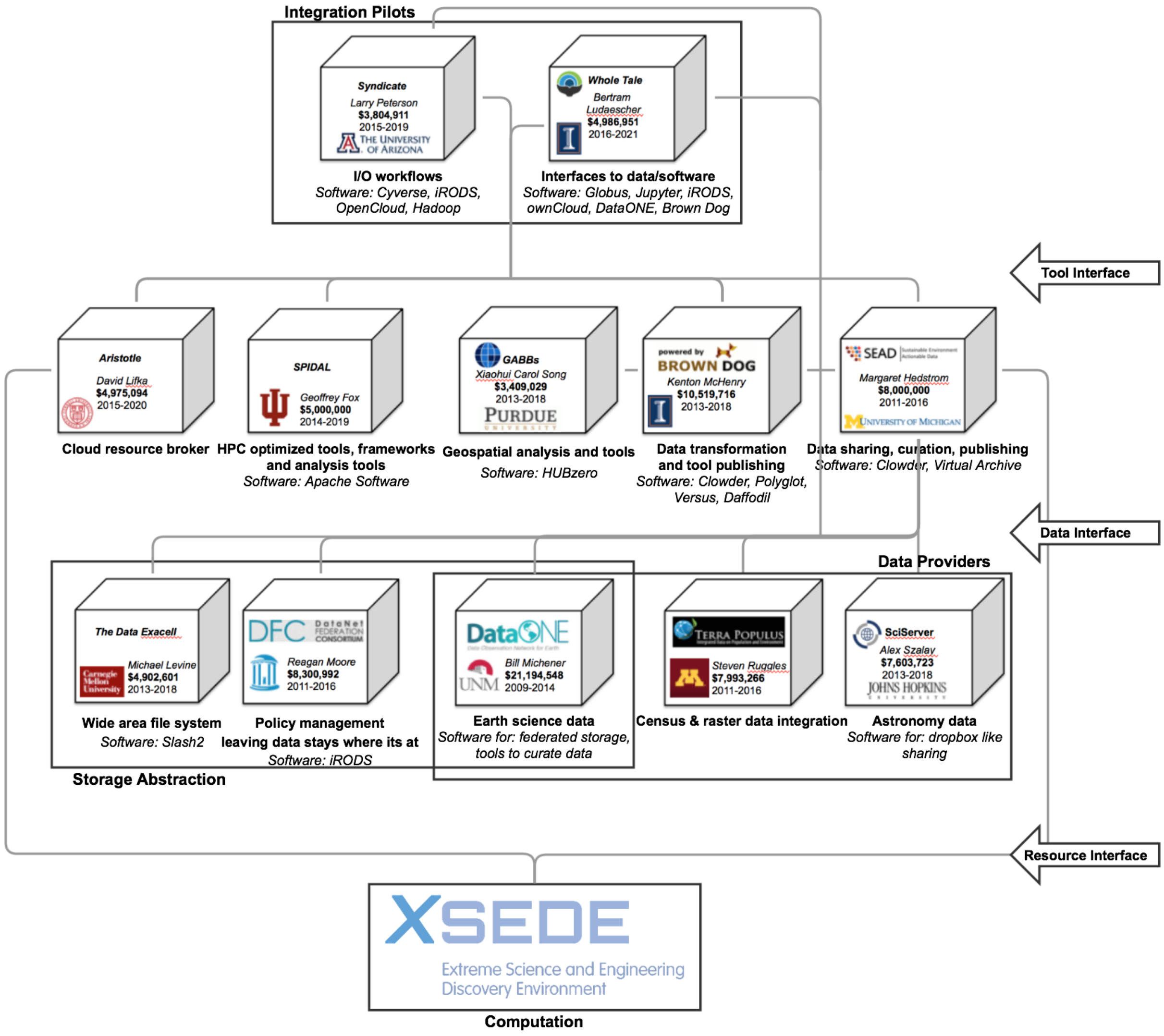


Figure 3: Interconnections Possible Between DIBBS and Datanet Projects.

1.4 NSF DIBBs Whole Tale

An example of a data management pipeline built on addressing interoperability among a number of major data management cyberinfrastructure components is planned within the DIBBs Whole Tale[18] project. Awarded in March 2016[19], the project aims to enable researchers to move, transform, examine, curate, and seamlessly publish research data artifacts as executable articles, enabling new discovery by making the synthesis of data easier for researchers. The project has begun with the drafting of a project execution plan and the identification of development staff.

1.5 NSF EarthCube

Through significant community outreach, landscape mapping, as well as investments in building cyberinfrastructure (CI) components of their own, the NSF EarthCube[20] effort has worked towards assembling

[18] http://wholetale.org/

[19] http://www.nsf.gov/awardsearch/showAward?AWD_ID=1541450

[20] http://earthcube.org/

an end-to-end data management pipeline for the geosciences. The NDS is coordinating with the EarthCube leadership towards leveraging one another's efforts, with members of EarthCube participating in NDS workshops and vice versa. There is interest in leveraging EarthCube's CINERGI[21] portal to catalog data management tools based on community input.[22] The EarthCube community could leverage the NDS Labs[23] resource towards increasing the exposure and usability of available data management tools within the geosciences. Both efforts plan to continue this mutual engagement towards the eventual establishment of a broadly usable scientific data management pipeline.

1.6 NSF Big Data Innovation Hubs

The NSF Big Data Hub program[24] aims to coordinate, at the regional scale, activities addressing regional scientific challenges involving data. These identified regional research areas or themes and their spoke activities, will coordinate closely with the Hub center to maximally leverage regional resources and activities and to facilitate transfer of innovation between multiple sectors including government, non-profits and industry. The NDS participates in a number of these activities, specifically in the West and Midwest, towards areas such as digital agriculture, smart cities, and personalized medicine, while emphasizing the leveraging and reuse of already existing software tools and services within each region (i.e., supporting communities to use these tools, possibly across scientific domains). The West[25] Big Data Hub and the Midwest[26] Big Data Hubs were awarded late last year. Since then, executive directors were identified to coordinate Hub activities and a first round of spoke and planning awards were made. In the Midwest, the NDSC worked to support a ring activity on Data, Tools & Services cutting across the regions spokes in the following ways:

- Compiled a draft list of regional resources prior the submission of letters of intent in order to facilitate the identification and potential reuse of resources within submitted spoke proposals[27].
- Participated in the Midwest Hub All Hands meeting and conducted a Data, Tools, & Services breakout session to gather ideas for new directions and coordinated among the four Hubs in order to support the use of these resources amongst the spokes within each region.
- Explored the coordination of data, tools, and service activities across the four hubs, towards leveraging each other's activities and resources in order to better support regional activities and minimize redundancy.
- Organized a Data, Tools & Services working group to further pursue the cataloging, exposure, and use of regional datasets, tools, and services in support of spoke activities.
- Utilized NDS resources and provided support staff for Hub activities such as a University of North Dakota Early Career Big Data Summit Hackathon[28] and a University of North Carolina's Automating Archive Policy Enforcement using Dataverse and iRODS workshop[29] at the 2016 IASSIST conference.

[21] http://cinergi.sdsc.edu/
[22] https://nationaldataservice.atlassian.net/wiki/display/NDSC/EarthCube+Technology+Components
[23] http://www.nationaldataservice.org/projects/labs.html
[24] https://www.nsf.gov/pubs/2015/nsf15562/nsf15562.htm
[25] http://westbigdatahub.org/
[26] http://midwestbigdatahub.org/
[27] https://docs.google.com/document/d/1n6K0IGQ25PKuVCoxPdCqXUuGt6M0Bf-CoxhPL7PVu4E/edit?pref=2&pli=1
[28] https://und.edu/research/computational-research-center/mbdh/early_career_big_data_summit_2016.cfm
[29] http://iassist2016.org/workshops.html

1.7 National Data Service Workshops

Since its inception, NDS has held two workshops each year towards bringing together members of the scientific, cyberinfrastructure, research library, and publishing communities. Workshop themes have varied: NDS 5 was held at the University of North Carolina at Chapel Hill, April 4th - 6th, 2016. This workshop highlighted the activities of the NSF DataNets and DIBBs awards, in particular the underlying technologies developed, and brought about discussion towards creating an end-to-end scientific data management pipeline. Representatives from nearly all of these activities were in attendance and presented a brief development overview and the use cases served. One of the breakout groups suggested that an Interoperability Interest Group be formed to examine which interfaces might be required to interconnect the DataNets and DIBBs technologies. In addition to the main program, the alpha version of the NDS Labs resource was demoed[30]. NDS 5 presentations are archived on the NDS website[31].

Picture 1: NDS5 Workshop

1.8 Interoperability Task Force

Interoperability among data management tools and services is a key element towards the realization of an end-to-end scientific data management pipeline that would make up a National Data Service. The TAC has outlined a plan[32] for the formation of an Interoperability Task Force towards fleshing out the interfaces, and in turn, the standards required to connect data management components. Once requirements are established, the task force will recommend interfaces that data management components might implement and work with pilot efforts to demonstrate such implementations.

2 Consortium Development Activities

Based partly on outcomes from the NDS workshops, a high level Work Breakdown Structure (WBS) was drafted outlining development activities for two resources over the next two years, NDS Labs and NDS

[30] https://www.youtube.com/channel/UCWuPo7LDCzsqF3RaKzIQmHw

[31] http://www.nationaldataservice.org/get_involved/events/NDS5/

[32] https://nationaldataservice.atlassian.net/wiki/display/NDSC/TAC+Interoperability+Task+Force

Share[33]. The draft WBS was reviewed by the TAC and development activities were initiated towards building up the first of the two resources, NDS Labs.

2.1 NDS Labs

Development activities focused on NDS Labs implementation as inspired by the 2nd and 3rd NDS workshops[34]. NDS Labs is an exposure and development resource meant to further the NDS mission of moving towards a national data service defined within its charter[35] by bringing together component technologies and allowing their development teams and users to work on the implementation of interoperability between them[36]. In addition to addressing component interoperability, the Labs resource also addresses needs within other efforts such as a community testbed resource for the RDA.

2.1.1 NDS Labs and the NDS Labs Workbench (alpha release)

The NDS Labs Workbench is an initial development phase aimed at bringing into focus a set of concrete development requirements for NDS Labs. A Platform as a Service (PaaS), NDS Labs allows for the easy deployment of currently active data management tools and services allowing users to experiment with them in a much more convenient manner than if they had to find, download, and install them themselves. Striving for maximal interoperability, NDS Labs allows users to connect different components and build new services with supports development between those components. The Labs environment enables users to easily add new tools and services to the catalog of services. These capabilities are intended to be as user friendly as possible to cater to the diverse scientific community skill levels. Mockups and storyboards for this initial phase are available on the NDS website[37].

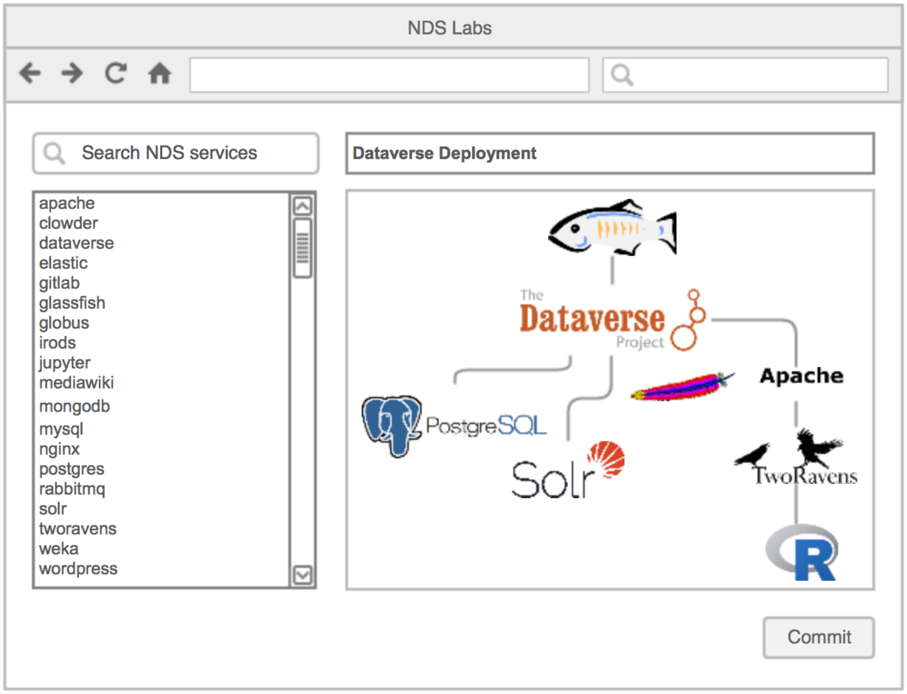


Figure 4. NDS Labs mockup

[33] https://nationaldataservice.atlassian.net/wiki/display/NDSC/Seed+Effort+Work+Breakdown+Structure
[34] http://www.nationaldataservice.org/docs/NDS3_Workshop_Report-v1.0.pdf
[35] http://www.nationaldataservice.org/docs/Charter_v2.pdf
[36] https://nationaldataservice.atlassian.net/wiki/display/NDSC/NDS+Labs
[37] https://nationaldataservice.atlassian.net/wiki/display/NDSC/NDS+Labs

Utilizing a graphical interface inspired by that of Juju charms[38] for deploying system components for the web, the **NDS Labs Workbench** interface contains two parts: a list of tools and services available for deployment on the left, and a canvas like area to the right where users can drag, drop, and connect components for deployment.

2.1.1.1 Services

An initial list of services that can be included as resources within the NDS Labs Workbench (see the left pane above) was drafted[39]. The list includes active data management tools being created by NSF, NIST, NIH, etc as well as the other open source component technologies which they build off of (e.g., databases such as mongodb, postgresql, messaging buses such as rabbitmq). From this list we identified services that the development team then worked to include as demonstrations of the Labs resource. We emphasize that a goal of the effort is to simplify the process of adding tools and services so that external parties can include their own. This is critical as we currently do not, nor may ever, have the resources to add the many heterogeneous data management resources currently being developed.

2.1.1.2 Developers API

An underlying Application Programming Interface (API)[40] was constructed to allow tech savvy contributors to directly interface with the available services within their own applications and to allow our own development of the front-end Workbench interface to be cleanly separated from underlying capabilities. Specifically, the API allows for Labs encapsulated tools and services to be provisioned and deployed programmatically. This includes authentication and allows for novel client applications to be built with the API.

2.1.1.3 Architecture

Building off of OpenStack[41] and the resources at TACC (*Rodeo)*, SDSC (*SDSC Cloud)*, and NCSA (*NCSA Nebula)*, an architecture based on the extension of Kubernetes[42] to deploy services and their components within a number of independent and interconnected docker[43] containers was chosen. Other leveraged technologies include etcd[44] for the backend of a catalog containing the available services and AngularJS[45] for the frontend. A high level diagram of the architecture is shown below:

[38] https://jujucharms.com/

[39] https://nationaldataservice.atlassian.net/wiki/display/NDSC/NDS+Labs+Services

[40] https://opensource.ncsa.illinois.edu/confluence/display/NDS/NDS+Labs+API

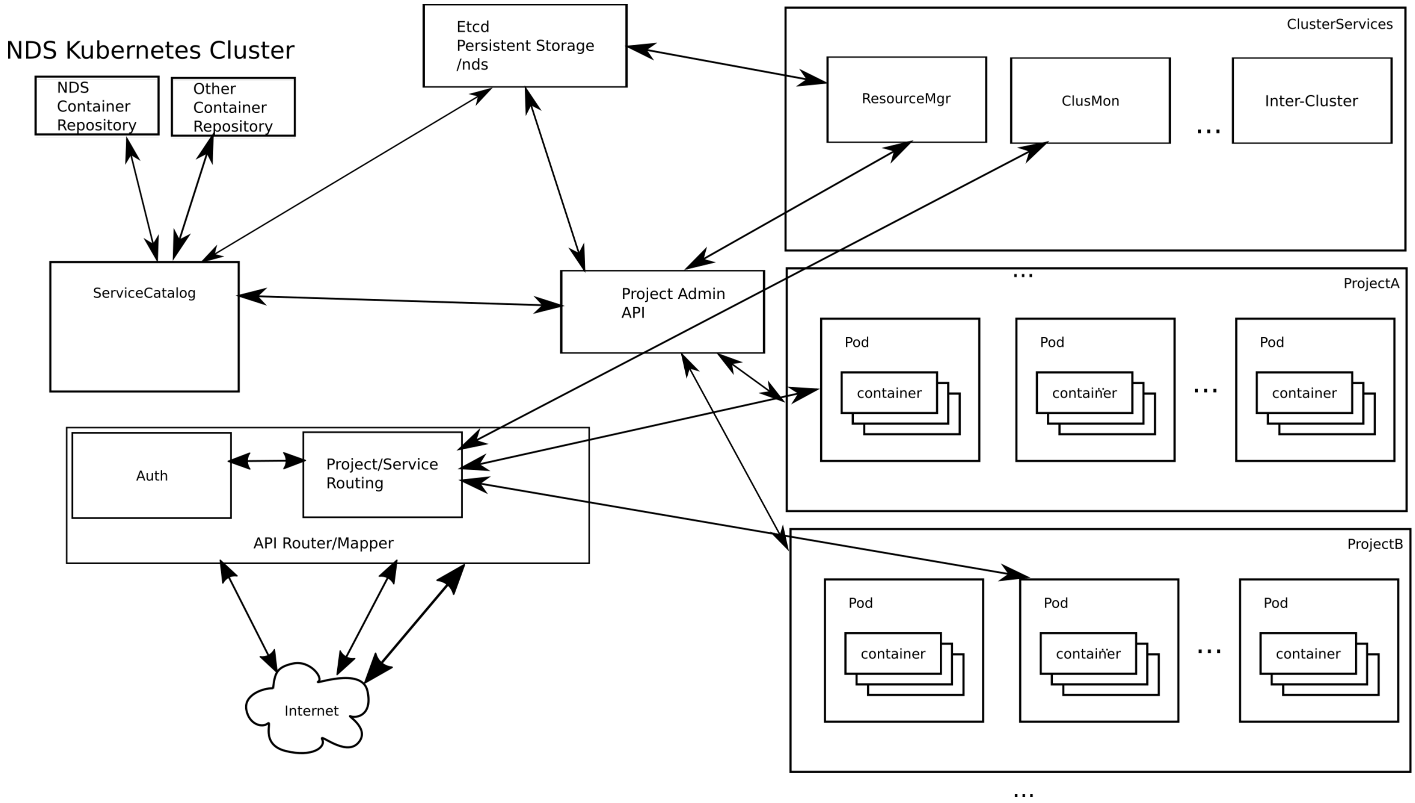
[41] https://www.openstack.org/

[42] http://kubernetes.io/

[43] https://www.docker.com/

[44] https://coreos.com/etcd/

[45] https://angularjs.org/



2.1.1.4 Development

The alpha release, unveiled at the NDSC5 Workshop in April of 2016, includes most of the architecture being implemented, the graphical interface, and a handful of included services that can be used as demonstrations within the workshop. Development activities can be followed directly in the Jira issue tracking system[46].

2.1.1.5 NDSC5 Preparations

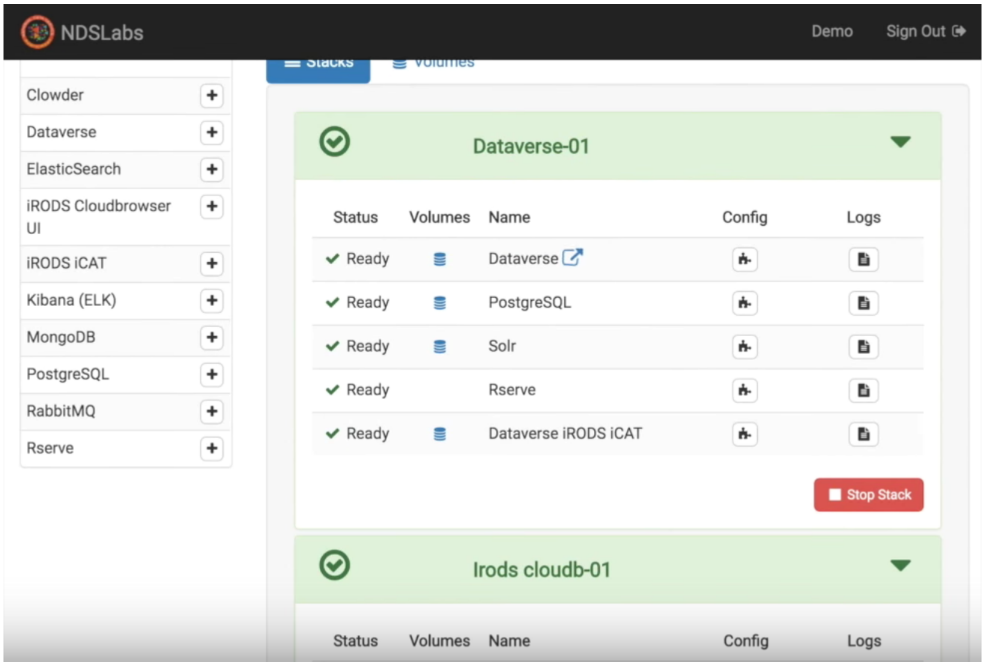
Preparations for presenting an alpha NDS Labs Workbench at NDSC5 workshop included creating demonstrations involving the deployment of the data management technologies used in two NDS collaborative efforts from their component technologies within the workbench (specifically the Odum Archive at UNC and the ARPAE TERRA-REF effort). Activities included:

- Working with the Odum and Terra teams to dockerize and reproduce their component technologies within the NDS Labs Workbench.
- Establishing an NDS YouTube channel and creating demonstration videos of the workbench: https://www.youtube.com/channel/UCWuPo7LDCzsqF3RaKzIQmHw
- Preparing tutorial materials and conducting a hands on tutorial showing both how to use the workbench as well as how one can add their own services to the workbench for others to use.
- Creating an NDS Labs google group to handle Q&A involved in the use and inclusion of services within the NDS Labs Workbench: https://groups.google.com/forum/#!forum/ndslabs

[46] https://opensource.ncsa.illinois.edu/jira/projects/NDS/

2.1.1.6 Alpha Release

The final look and feel for the alpha workbench is shown below. Users can select data management tools and services from the left pane, add them with their dependencies on the right, configure the service, examine its logs, and interact with the service if it provides a web interface.



Dataverse-01

Status	Volumes	Name	Config	Logs
✓ Ready	●	Dataverse		
✓ Ready	●	PostgreSQL		
✓ Ready	●	Solr		
✓ Ready		Rserve		
✓ Ready	●	Dataverse iRODS iCAT		

Irods cloudb-01

Status	Volumes	Name	Config	Logs

While the main purpose of the NDS Labs Workbench is to foster the development of interoperability between active data management components, which will be addressed in the next development phase for a beta release of the Workbench, initial user feedback has shown that the Workbench, even at its alpha state, fills a scientific community need in allowing projects to quickly identify and try several alternative tools and services for their projects data management needs. This is otherwise currently difficult to do, as there are many data management technologies under development, many of which are isolated in pockets of communities, and many of which have varying degrees of difficulty in terms of getting access to and/or standing up in order to try out.

2.1.2 NDS Labs Workbench IASSIST Workshop Milestone

In preparation for the upcoming UNC Odum Institute workshop at IASSIST16[47] the development team began work to deploy the Workbench as a web service, whereas previously it was run locally per user. Provisioned with the data management technologies that would be covered at the workshop the Workbench would provide a convenient means by which attendees could signup, deploy, and play with these data management tools without having to install them themselves. Development for this milestone addressed resource provisioning, authentication, as well as the hardening of user roles.

2.1.3 NDS Labs and NDS Labs Workbench (beta release)

The beta release, planned for the NDSC6 workshop (Fall 2017), focuses on fleshing out the development features and components of the NDS Labs Workbench that will directly address fostering interoperability between actively developed data management technologies. Planning of development activities can be

[47] http://iassist2016.org/

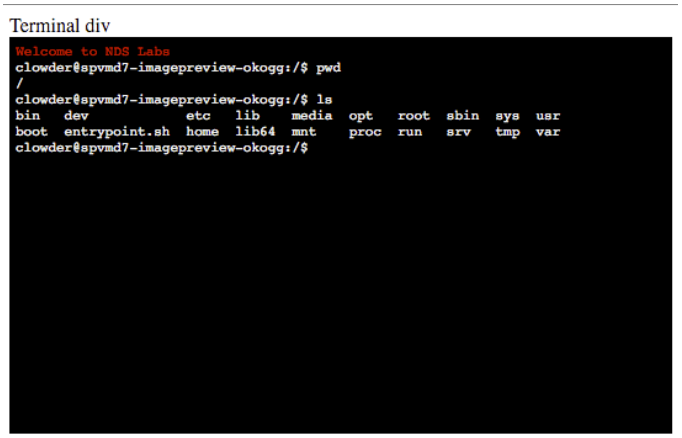
followed on the NDS wiki[48], working around a scenario as follows: a developer wishes to implement a tool that enables a federated search across the NIST Materials Data Facility[49] (built on Globus Publish[50]) and the DIBBs T2C2 4CeeD collection (built on DataNet SEAD[51] and Clowder[52]). We will enable and enhance two possible approaches addressing this scenario within the NDS Labs environment:

1. The developer creates a **new** tool or service that searches across these collections. This can take the form of a tool that is aware of the APIs for both Globus and Clowder and internally uses them to pass along queries to both collections. The NDS Labs Workbench can foster this scenario by providing easily deployable sample instances of the two underlying services, allowing the developer to work and debug against them in a controlled, accessible, and isolated development environment.

2. The developer **extends** the two underlying technologies of the two collections so that an existing tool or service can query them. This can take the form of the developer branching and modifying Clowder to add support for a standard such as OAI-PMH[53], an emerging standard for accessing metadata within archives, and then doing the same in Globus Publish. Any search tool capable of searching across archives supporting OAI-PMH can then be used. The NDS Labs Workbench can foster this by providing pre-configured development environments for the two technologies and allowing the user to easily try out/deploy their changes within NDS Labs and possibly test them with an available search tool contained in the catalog of tools and services.

Development of the NDS Labs Workbench will incorporate features and resources such as the following, to enabling these developer scenarios:

- In addition to buttons allowing the user to access the service logs, configuration, and web interface if available, the NDS Labs Workbench will add a button to allow developers to access a terminal to the container running the service (useful in allowing a developer to interact with the service at a more direct lower level). This feature has been implemented by leveraging term.js[54] to provide a web based terminal.



Terminal div

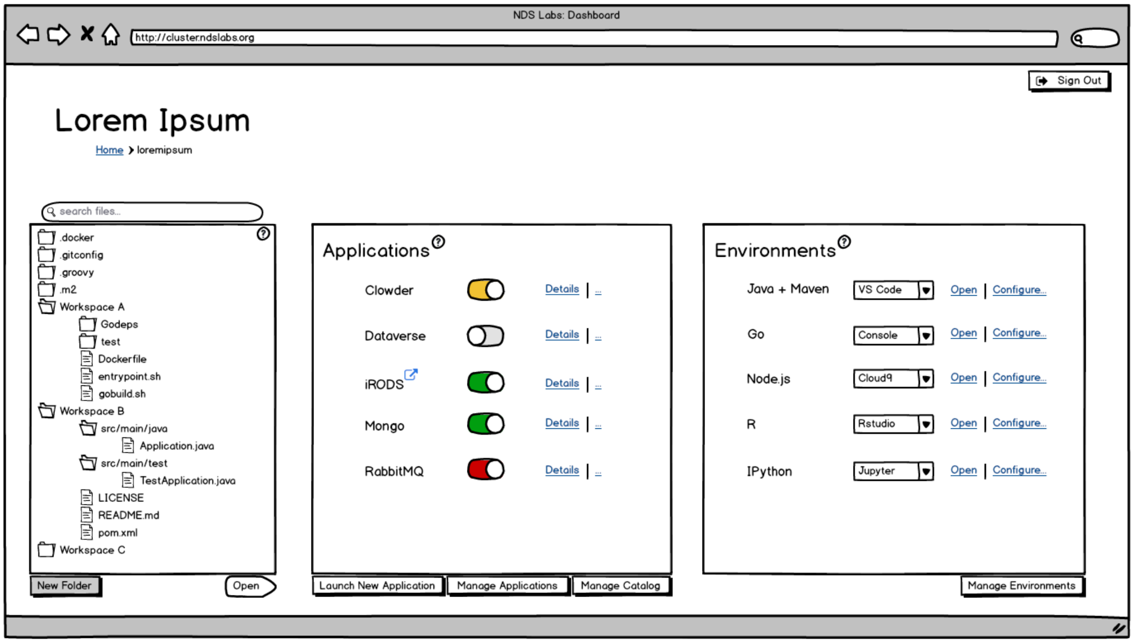
```
Welcome to NDS Labs
clowder@spvmd7-imagepreview-okogg:/$ pwd
/
clowder@spvmd7-imagepreview-okogg:/$ ls
bin   dev            etc   lib    media  opt   root  sbin  sys  usr
boot  entrypoint.sh  home  lib64  mnt    proc  run   srv   tmp  var
clowder@spvmd7-imagepreview-okogg:/$
```

[48] https://opensource.ncsa.illinois.edu/confluence/display/NDS/Developer+support+in+NDS+Labs
[49] https://www.materialsdatafacility.org/
[50] https://www.globus.org/
[51] http://sead-data.net/
[52] https://clowder.ncsa.illinois.edu/
[53] https://www.openarchives.org/pmh/
[54] https://github.com/chjj/term.js

- Containerizing development environments associated with the tools and services include in the Labs Workbench. By doing this we can alleviate a significant amount of effort required as a first step by new developers to a project. These development environments would be different for each tool, including needed dependencies such as languages, libraries, and other software utilized to obtain, edit, and launch the relevant code. Generic development environments will also be provided (e.g., for general Python or Java development).
- Integration with relevant service repositories allowing developers to quickly identify needed code repositories and branch off of them to store changes.
- Integration with distributed file system technologies (e.g., Samba, NFS, FTP, SFTP, DropBox, ownCloud[55]) allowing developers to leverage the web-based IDE or their own perhaps more familiar local development tools, potentially switching back and forth, assuming they locally connect to or host the distributed storage.
- Incorporating a web-based Integrated Development Environment (IDE) allowing for code development to occur directly from within the online Labs Workbench environment. Made popular initial in web development, e.g., JavaScript development, web-based IDEs such as Codenvy[56], Cloud9[57], noVNC[58], Xpra[59], etc, now support a wide variety of programming languages such as Python, Java, and C++.
- Allowing for the rapid deployment of services, possibly from within the IDE, for testing and debugging during development. This could take the form of a deploy button alongside the development environment that takes the current state of the service and any modified code and deploys it within Labs.

The Labs Workbench interface will also be modified to better organize these capabilities, allowing both novice users and more technical users to easily use and navigate the resources provided. When a user first logs in, they will be presented with an “App Store” like view highlighting the available data management tools and services at their disposal. From here, they can either deploy tools or services, or launch associated development environments to allow interaction, and modification of options. A subsequent login session will bring the user to a dashboard, with a mockup (shown below), allowing the user to see currently deployed services within their space, resources being utilized, etc.

[55] https://owncloud.org/
[56] https://codenvy.com/
[57] https://c9.io/
[58] https://kanaka.github.io/noVNC/
[59] https://xpra.org/



2.2 NDS Share

The second resource planned for development by the NDS is NDS Share. NDS Share will allow for the deployment and/or integration of external production instances of data management tools and services, allowing a member of the scientific community to both more quickly identify a repository for produced digital data artifacts as well as find and obtain datasets across a variety of archives and fields of study. Development of NDS Share is in its very initial phases and has thus far included the activities outlined below.

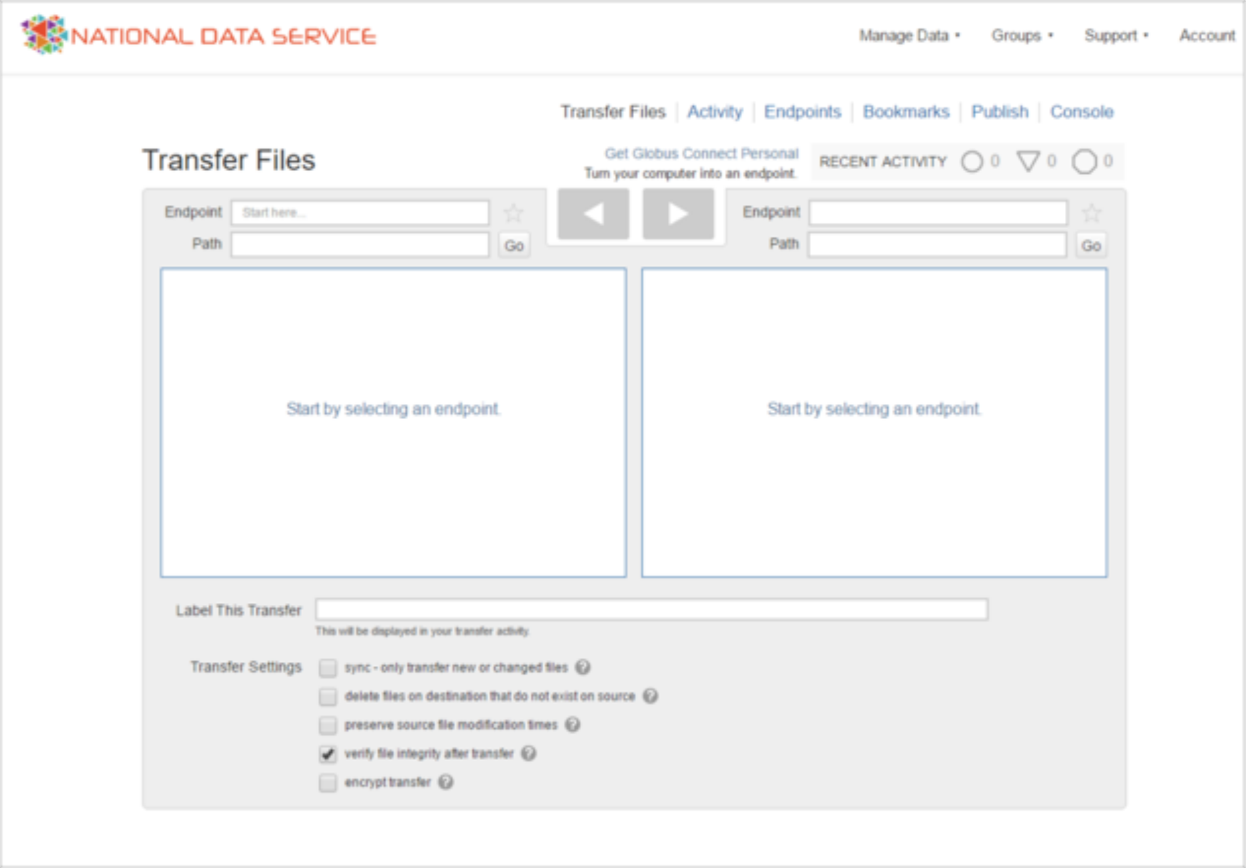
2.2.1 Planning

The TAC has drafted an overview document that outlines NDS Share's offerings and capabilities[60]. This initial document is being reviewed and modified by the TAC as well as the NDS Executive Committee. Once completed, the NDS Technical Coordinator and development team will utilize this to flesh out a design along with development and feature requirements.

2.2.2 Repository of Last Resort

A long running requirement, even in the discussions thus far for NDS Share, has been that of establishing an NDS-supported repository of last resort for those within the scientific community that are unable to find another location to publish and store data that they generate. This repository of last resort would include both the underlying storage as well as a deployed instance of one of the data management technologies on top of this resource allowing a user to move, manage, and publish data (going as far as obtaining a unique Digital Object Identifier (DOI) which can then be used to cite the data). The Globus team in collaboration with NCSA stood up an example repository of last resort, leveraging the Santiago storage resource at NCSA and Globus Publish as the data management technology: (http://data.share.nationaldataservice.org):

[60] https://nationaldataservice.atlassian.net/wiki/display/NDSC/NDS+Share


NATIONAL DATA SERVICE
Manage Data ▾
Groups ▾
Support ▾
Account
Transfer Files | Activity | Endpoints | Bookmarks | Publish | Console
Transfer Files
Get Globus Connect Personal
Turn your computer into an endpoint.
RECENT ACTIVITY ○ 0 ▽ 0 ○ 0
Endpoint
Start here...
Path
Go
Endpoint
Path
Go
Start by selecting an endpoint.
Start by selecting an endpoint.

Label This Transfer

This will be displayed in your transfer activity

Transfer Settings

- ☐ sync - only transfer new or changed files
- ☐ delete files on destination that do not exist on source
- ☐ preserve source file modification times
- ☑ verify file integrity after transfer
- ☐ encrypt transfer

This instance of a repository of last resort was presented at the NDSC5 workshop, but will not be made available for broad use until a longer term backend resource can be identified. NCSA is actively exploring establishing such a resource to provide a long-term storage location for highly valued datasets.

2.2.3 Repository Recommender Service

Another long running requirement, especially where a repository of last resort is involved, has been the need to provide a repository recommender service, that given information about a data set, will attempt to identify and recommend one or more institutional repositories that can be utilized for storing the data. Towards standing up such a service, the NDSC development team has begun looking at relevant resources to leverage, such as re3data.org[61], which has made a significant effort to catalog available archives as well as the interfaces they provide, and the DataNET SEAD Virtual Archive[62], which moves data from its Active Content Repository (ACR) to be published in a long term archive once collaboration and work on the data has been finalized.

3 Development Pilots and Collaborative Activities

The NDS has undertaken a number of collaborative efforts and sponsored a number of pilot projects to develop components and build interoperability amongst components, which may support a U.S. National Data Service. Below we describe these activities.

[61] http://www.re3data.org/

[62] https://github.com/Data-to-Insight-Center/sead2

3.1 NIST Materials Data Facility

The Materials Data Facility[63,64] (MDF) is a collaboration between Globus (University of Chicago), the National Center for Supercomputing Applications (NCSA-UIUC), and the Center for Hierarchical Materials Design (CHiMaD), a NIST-funded center of excellence. MDF is developing key data services for materials researchers with the goal of promoting open data sharing, simplifying data publication and curation workflows, encouraging data reuse, and providing powerful data discovery interfaces for data of all sizes and sources. Specifically, MDF services will allow individual researchers and institutions to 1) enable publication of large research datasets with flexible policies; 2) grant the ability to publish data directly from local storage without third-party publishers; 3) build extensible domain-specific metadata and automated metadata ingestion scripts for key data types; 4) develop publication workflows; 5) register a variety of resources for broader community discovery; and 6) access a discovery model that allows researchers to search, interrogate, and build upon existing published data.

As an NDSC collaborative effort, the MDF project is piloting a portion of a data management pipeline for the material sciences addressing data acquisition, data access control, data movement, curation, and publication. The underlying Globus[65] services further serve as an example of a non-domain specific general purpose service with a programmable REST interface allowing other tools to connect and build on top of it. Having stood up the MDF repository utilizing Globus and storage at the NCSA, the team works to engage the materials science community, exposing new researchers to the notion of data publication and credit, while also gathering more scientific collections that would otherwise be lost. In addition, the team, in collaboration with NIST, has leveraged the NDS Labs resource to deploy community resource registry services such as the Materials Resource Registry[66], allowing researchers to more easily identify relevant tools and data sources, and the International Metrology Resource Registry[67] for metrology resources.

3.2 ARPA-E TERRA-REF

Phenotypes are measurable features that indicate how an entity (such as a plant) will grow and respond to stresses such as heat, drought, and pathogens. Crop breeding is currently limited by the speed at which phenotypes can be measured, and the information that can be extracted from these measurements. Currently, measurements used to predict yield include measuring leaf thickness with a caliper or height with a meter stick. More sophisticated instruments used to quantify plant architecture, carbon uptake, water use, and root growth do not scale to the thousands or tens of thousands of individual plants that need to be evaluated in a breeding program. TERRA-REF[68,69] will develop an integrated phenotyping system for Sorghum that leverages genetics and breeding, automation, remote plant sensing, genomics, and computational analytics.

As an NDSC collaborative effort, TERRA-REF leverages and interconnects a number of data management technologies to implement a data management pipeline for the acquisition, movement, analysis, curation, and sharing of biological sensor data. Leveraged technologies include Clowder[70], which drives the ACR in DataNet

[63] https://materialsdatafacility.org/

[64] http://link.springer.com/article/10.1007/s11837-016-2001-3

[65] https://www.globus.org/

[66] http://matsci.registry.nationaldataservice.org/

[67] http://nist.registry.nationaldataservice.org/

[68] http://terraref.org/

[69] https://github.com/terraref

[70] https://clowder.ncsa.illinois.edu/

SEAD[71], CyberGIS[72], Globus, CyVerse[73], and tools such as PlantCV[74]. A further NDSC topic of interest that the TERRA-REF project will explore is search mechanisms that span several heterogeneous databases and/or repositories.

3.3 iSEE Plants in silico

As the Earth's population climbs toward 9 billion by 2050 — and the world climate continues to change, affecting temperatures, weather patterns, water supply, and even the seasons — future food security has become a grand world challenge. Accurate prediction of how food crops react to climate change will play a critical role in ensuring food security. The ability to computationally mimic the growth, development and response of plants to the environment will allow researchers to conduct many more experiments than can realistically be achieved in the field. Designing more sustainable crops to increase productivity depends on complex interactions between genetics, environment, and ecosystem. Therefore, creation of an *in silico* — computer simulation — platform that can link models across different biological scales, from cell to ecosystem level, has the potential to provide more accurate simulations of plant response to the environment than any single model could alone.

As an NDSC collaborative effort, Psi[75,76] aims to bring together a framework for the integration of various models and data sources that will be needed to simulate whole plants and/or crops. These models will span scales ranging from the molecular level, gene level, and phenotype level and require the linking of heterogeneous code and data requiring transformations across cluster/HPC resources. The team is exploring how they may leverage frameworks such as Cactus[77] or Swift[78] and tools such as Jupyter[79] notebooks towards building a framework for plant models supporting technical users and users within an educational setting. The effort also takes on a Software Carpentry[80] like role with NDSC staff helping the team organize and share their code, adopt good software practices, and gain exposure with new technologies towards connecting their models. Preliminary work was shown at the Plants in silico symposium on May 18th-19th 2016[81].

3.4 NIH BD2K KnowEng

KnowEnG[82,83] (pronounced "knowing") is a National Institutes of Health-funded initiative that brings together researchers from the University of Illinois and the Mayo Clinic to create a Center of Excellence in Big Data Computing. It is part of the Big Data to Knowledge (BD2K) Initiative that NIH launched in 2012 to tap the wealth of information contained in biomedical Big Data. KnowEnG is one of 11 Centers of Excellence in Big Data Computing funded by NIH in 2014. This four-year project will create a platform where biomedical scientists, clinical researchers, and bioinformaticians can bring their own data and perform common as well as

[71] http://sead-data.net/

[72] http://cybergis.illinois.edu/

[73] http://www.cyverse.org/

[74] http://plantcv.danforthcenter.org/

[75] http://sustainability.illinois.edu/research/climate-solutions/plants-in-silico-project/

[76] https://github.com/rachelshekar/Plants_in_Silico

[77] http://cactuscode.org/

[78] http://swift-lang.org/

[79] http://jupyter.org/

[80] http://software-carpentry.org/

[81] http://sustainability.illinois.edu/outreach/plants-in-silico-conference/

[82] http://www.knoweng.org/

[83] https://github.com/KnowEnG-Research

advanced analysis tasks, guided by the “knowledge network”, a large compendium of public-domain data. The knowledge network embodies community data on genes, proteins, functions, species, and phenotypes, and relationships among them. Instead of analyzing their data set in an isolated fashion, researchers will be able to go straight to asking global questions. The infrastructure, capacity and tools will grow with the datasets.

As an NDSC collaborative effort, KnowEng develops services for the integration and analysis of data, specifically genomics data. Stored as a knowledge network with genes at the nodes and relationships amongst the genes, aggregated from multiple database sources, as edges, the team works to develop tools to leverage this union of information in order to answer a variety of novel questions. Leveraged data management tools include HUBzero[84], with additional tools for clustering, regression, classification, and set characterization being developed.

3.5 Renaissance Simulations

Using the powerful visualization and analysis package, yt[85], as an exemplar, this project is creating flexible and reusable recipes for creating presentations of data customized for a particular community. Going beyond the simple splash page, this project leverages cloud technologies for putting advanced interfaces in front of data. In particular, it enables scientists to safely apply custom analysis to remote data in the form of, for example, Python scripts.

This pilot effort is utilizing NDS Labs resources to host its archive of simulations. NDS staff run a specialized server where an NDS-inspired set of tools allows users to view Jupyter notebooks, run analysis in Dockerized containers and to add their own findings in additional Jupyter notebooks.

3.6 Data Access Alliance with Globus and Dataverse

In support of the Structural Biology Grid Data effort (SBGrid), the NDS team is providing storage space on SDSC project storage to complement stores at the Harvard Medical School and NCSA. As part of this pilot effort we are developing new software techniques to algorithmically move data towards compute allocations as well as to move data by (usage) heat. This effort includes adding support to Globus to interface with object storage resources which once completed would become a valued feature to the Globus community.

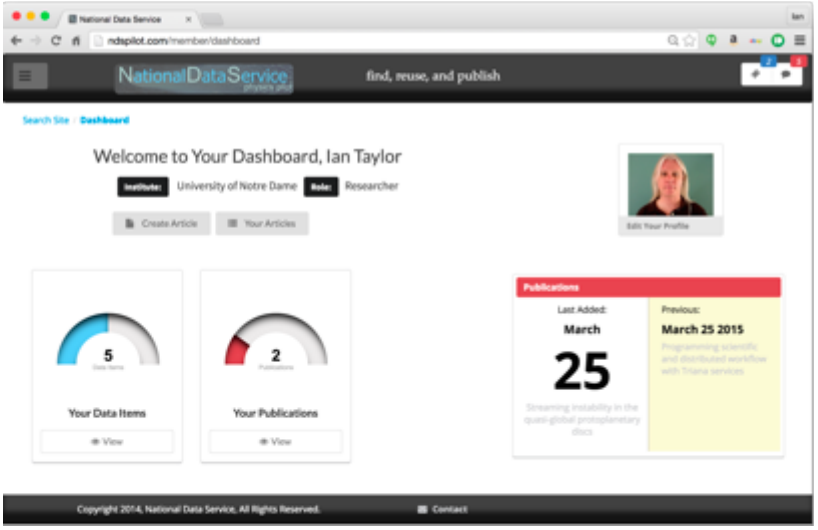
3.7 NDS Researcher Dashboard Pilot

The University of Notre Dame and Cardiff University, UK, have been working on a dashboard concept that enables researchers a means of interacting with existing research[86]. This work was built in collaboration with the NDS and aims to provide an intuitive Web-based interface to expose fully interactive research containers that support the lifecycle of scholarly communication. Research containers enable executable and repeatable research by supporting methods, source code, and data within dynamically created Docker containers. To date, there have been two versions of the dashboard.

[84] https://hubzero.org/

[85] http://yt-project.org/

[86] https://bitbucket.org/nds-org/nds-dashboard



Welcome to Your Dashboard, Ian Taylor

Institute: University of Notre Dame Role: Researcher

Publications

Last Added: March 25

Streaming instability in the quasi-global protoplanetary discs

Previous: March 25 2015

Programming scientific and distributed workflow with Triana services

Copyright 2014, National Data Service. All Rights Reserved.

The V1.0 dashboard integrated with the NDS Labs *Epiphyte* API, which interfaced with Docker containers and data management systems. This dashboard used Yii and an SQL engine that was modelled using the OAI-ORE Web aggregation standard to allow interoperability with other repositories. The system has been demonstrated at NDS meetings (Washington DC and Austin), and at SC14 at the NCSA booth. The Dashboard V2 interfaced to the Open Science Framework (OSF)[87], which is an environment that supports open materials, data, tools to connect projects and initiatives and easy online publishing of results. Using this system, a researcher can create a project on the OSF, connect data management tools to it (e.g. Google Drive, Dropbox, Box, Dataverse, etc.) and then use the dashboard to execute methods on data, stored on OSF, using the Boatload API[88], an API for automating deployment and operations of Docker containers on clusters.

3.8 Adaptations to Existing National Data Infrastructure Technologies

NDSC represents a growing consortium of contributing researchers and service providers. Many of these contributing providers have undertaken work to enhance their services to better serve the needs of NDSC users. Globus is one example of a foundational service that provides general research data management capabilities. Globus provides high performance access to data using the GridFTP protocol as well as foundational identity and access management and is deployed at hundreds of institutions across the country. It is therefore integral to the vision of NDS as it enables robust, high performance access to, and management of distributed data. Over the period of this report Globus has made several enhancements that are specifically relevant to NDSC. Specifically, it now includes a new extensible authentication and authorization model, software to support access to data stored on an increasingly broad set of storage systems, a secure HTTPS model that enables direct access to data, and a prototype secure and scalable data search services. These enhancements are in addition to the data transfer, sharing, and publication capabilities that are used by many NDS users and participant services.

NDSC participants represent a diverse set of users and providers with huge amounts of data distributed over many different types of storage systems. Increasingly these storage systems are built upon new storage architectures (e.g., object storage or distributed file systems). To better address the changing storage landscape Globus is actively developing new support for different classes of storage. Globus now supports (or

[87] http://osf.io

[88] https://bitbucket.org/keyz182/boatload

will soon support) Linux, Windows and MacOS file systems, Lustre, GPFS, HDFS, HPSS, Spectra BlackPearl, Amazon S3, OpenStack Ceph and Google Drive. This flexible storage support enables NDSC participants to access data using a standard Globus interface irrespective of the physical storage architecture used.

Given the enormous amount of data accessible via Globus and in other data repositories, there is a growing need for better methods to search and discover data. The Globus team have developed a prototype search service that aims to ingest and index metadata related to data residing in distributed Globus endpoints and other data repositories. At the heart of this model is a standard, yet flexible, model for representing metadata and a scalable and secure free-text search service. Importantly the model is equipped to enforce dynamic access permissions on every item indexed, thus different users will see different results for the same search query. The initial implementation is currently being piloted via collaborations with various users, including the MDF effort mentioned previously. It has been used to index a range of publicly accessible Globus data as well as publicly accessible data from several domain-specific and institution data repositories.

4 Personnel

Utilizing seed funding from NCSA and leveraging the organization's new Integrated Cyberinfrastructure directorate, development staff and activities on NDS Labs and Share were started based on requirements defined at the 2nd and 3rd NDSC Workshops. Identified personnel to date include:

- **Kenton McHenry**, Deputy Director of NCSA's Scientific Software & Applications division, named *NDS Technical Coordinator*
- **Mike Lambert**, *Research Programmer*, with a strong background in web development Mike contributes towards front end and user interaction elements of the NDS Labs Workbench.
- **Craig Willis**, *Research Programmer*, with a library science background and a focus querying data collections, Craig contributes towards backend development, service/tool incorporation, as well as community interaction.
- **David Raila**, *Research Programmer*, with a systems background David contributes to backend development for the NDS Labs Workbench and supports the iSEE Plant in silico effort.
- **Michal Ondrejcek**, *Research Programmer*, Ph.D. in Material Science, developer and support for community engagement for the NIST Materials Data Facility led out of Argonne.
- **Rob Kooper**, *Senior Research Programmer*, coordinates development activities on the ARPAE Terra effort, initially supported by Dora Cai, Senior Database Architect of the Application's Support team, and long term by Craig Willis who will explore federated search aspects within the project.
- **Jing Ge**, *Research Programmer*, has begun work on the NIH BD2K KnowEng effort.
- The recruitment of the *NDS Executive Director*, **Christine Kirkpatrick**, Division Director of IT Systems & Services at the San Diego Supercomputer Center joining NDS as of July 1st.

Executive Committee

John Towns
(jtowns@illinois.edu)
NDS Director

Executive Director,
NCSA
University of Illinois at
Urbana-Champaign

Ian Foster

Project Director, Globus
University of
Chicago/Argonne
National Labs

Niall Gaffney

Director of Data
Intensive Computing,
TACC
University of Texas,
Austin

Heidi Imker

Director, UI Research
Data Service
University of Illinois
Urbana-Champaign

Michael Norman

Director, SDSC
University of California
San Diego

Ed Seidel (Chair)

Director, NCSA
University of Illinois
Urbana-Champaign

Steering Committee

Alan Blatecky

RTI International,
Visiting Fellow

Mercè Crosas

Chief Data Science and
Technology Officer
IQSS
Harvard University

Ted Habermann

Director, Earth Science
The HDF Group

Robert J. Hanisch
(Chair)
Director, Office of Data
and Informatics
Material Measurement
Laboratory
National Institute of
Standards and
Technology

Carole Palmer

Associate Dean,
Research Information
School
University of
Washington

Ed Seidel

Director, NCSA
University of Illinois
Urbana-Champaign

Anita de Waard

VP, Research Data
Services
Elsevier Publishing

John Towns
(jtowns@illinois.edu)
NDS Director

Executive Director,
NCSA
University of Illinois at
Urbana-Champaign

Technical Advisory Committee

Christine Kirkpatrick

SDSC
University of California
San Diego

Kyle Chard

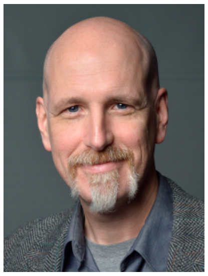
Argonne National Labs
University of Chicago

James Myers

School of Information
University of Michigan

Jarek Nabrzyski

University of Notre
Dame



Ray Plante

National Institute for
Standards and
Technology

Matt Turk

Research Scientist,
NCSA
University of Illinois at
Urbana-Champaign

Project Office

John Towns
NDS Director

Executive Director, NCSA
University of Illinois
at Urbana-Champaign

Kenton McHenry
NDS Technical Coordinator

Deputy Director, Scientific Software &
Applications Division, NCSA
University of Illinois
at Urbana-Champaign

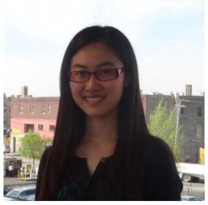
Kandace Turner
NDS Project Manager

Project Manager, NCSA
University of Illinois
at Urbana-Champaign

Development Team

Ben Blaiszik

Software Developer,
Computation Institute
University of Chicago



Jing Ge

Research Programmer, NCSA
University of Illinois
at Urbana-Champaign

Rob Kooper

Senior Research Programmer, NCSA
University of Illinois
at Urbana-Champaign

Mike Lambert

Research Programmer, NCSA
University of Illinois
at Urbana-Champaign

Michal Ondrejcek

Research Programmer, NCSA
University of Illinois
at Urbana-Champaign

David Raila

Research Programmer, NCSA
University of Illinois
at Urbana-Champaign

Ian Taylor

Adjunct Research Associate
Professor, CRC
Notre Dame

Craig Willis

Research Programmer, NCSA
University of Illinois
at Urbana-Champaign

5 Community Engagement

Engagement with researchers, infrastructure developers, libraries, and publishers is a significant part of the activities carried about by the NDSC towards realizing a national data service. Activities to date:

- Anita DeWard, Bob Hanisch, Christine Kirkpatrick, Ray Plante and Craig Willis attended the 8th RDA Plenary, Denver, September 2016, https://rd-alliance.org/plenaries/rda-eighth-plenary-meeting-denver-co/
- David Raila attended the Dataverse Community Meeting, Harvard, July 2016, http://projects.iq.harvard.edu/dcm2016/home
- Christine Kirkpatrick presented to the Hub-wide Data Sharing and Infrastructure meeting hosted by the South Big Data Innovation Hub, virtual meeting, July 2016
- Christine Kirkpatrick attended Data Commons and Data Sharing Workshop, Chicago, June 2016, https://sites.google.com/site/datacommons2016/
- Christine Kirkpatrick attended 2nd Nordic Data Services Workshop, Sweden, May 2016, https://wiki.neic.no/wiki/2nd_Nordic_data_services_workshop
- Kenton McHenry was invited to participate in the EarthCube Architecture Workshop, San Diego, May 2016, http://earthcube.org/workspace/technology-architecture-committee/architecture-workshop-may-2016
- Kenton McHenry presented a keynote at the Workshop on Container Strategies for Data & Software Preservation, Notre Dame, May 2016, https://daspos.crc.nd.edu/index.php/workshops/container-strategies-for-data-software-preservation-that-promote-open-science
- Kenton McHenry presented at Plants in Silico Symposium, Urbana, May 2016, http://sustainability.illinois.edu/outreach/plants-in-silico-conference/
- Kenton McHenry presented at Best Practices in Data Infrastructure Workshop, Pittsburgh, May 2016, https://www.psc.edu/index.php/bpdi-workshop
- Kenton McHenry participated in RDA/US Leadership Meeting, Albany, May 2016
- Craig Willis presented at Harvard Medical School, Harvard, May 2016
- David Raila presented at NCSA Industry Annual Meeting, Urbana, May 2016, http://www.ncsa.illinois.edu/Conferences/2016Meeting/agenda.html
- Kenton McHenry presented at CHiMAD workshop on Building an Interoperable Materials Data Infrastructure, Evanston, May 2016, http://chimad.northwestern.edu/news-events/Event_Archives.html
- John Towns and Craig Willis presented at DIBBs T2C2: Automated Cyber-Environments for Semiconductor Fabrication Data Collection and Analysis (CyberFab) Workshop, Urbana, May 2016, http://t2c2.csl.illinois.edu/workshop/
- Christine Kirkpatrick and Sharief Youssef presented at the West Big Data Hub All Hands Meeting's Collaboratory Faire, Berkeley, May 2016, http://westbigdatahub.org/activities/all-hands-2016/
- Kenton McHenry presented at Computer Science IT group meeting, Urbana, April 2015
- 5th National Data Service Consortium Workshop, North Carolina, April 2016, http://www.nationaldataservice.org/get_involved/events/NDS5/
- Christine Kirkpatrick, Kenton McHenry and Ed Seidel presented at Midwest BigData Hub All Hands Meeting, Chicago, March 2016, http://midwestbigdatahub.org/
- Anita DeWard, Kenton McHenry, Bob Hanish, Ray Plante, Carole Palmer presented at Research Data Alliance 7th Plenary Meeting, Tokyo, February 2016, https://rd-alliance.org/plenaries/rda-seventh-plenary-meeting-tokyo-japan
- Ray Plante, Ben Blaiszik, and Kenton McHenry attended NIST/CHiMAD Data, Databases, & Discovery workshop, Evanston, January 2016, http://chimad.northwestern.edu/news-events/past-events.html
- John Towns and Kenton McHenry presented at NCSA International Partners Meeting, IEEE Supercomputing, Austin, November 2015, http://sc15.supercomputing.org/

- Kyle Chard, Ian Foster, David Raila, Matt Turk, Kacper Kowalik, and Kenton McHenry demoed software and tools at IEEE Supercomputing, Austin, November 2015, http://sc15.supercomputing.org/
- Ray Plante and David Raila participated in EUDAT Working Groups Workshop, Barcelona, November 2015, https://www.eudat.eu/events/eudat-first-working-groups-workshop-12-13-november-2015-barcelona-spain
- 4th National Data Service Consortium Workshop, San Diego, October 2015, http://www.nationaldataservice.org/get_involved/events/NDS4/
- 3rd National Data Service Consortium Workshop, Austin, March 2015, http://www.nationaldataservice.org/get_involved/events/NDS3/
- 2nd National Data Service Consortium Workshop, Washington DC, October 2014, http://www.nationaldataservice.org/get_involved/events/NDS2/
- 1st National Data Service Consortium Workshop, Boulder, June 2014, http://www.nationaldataservice.org/get_involved/events/NDS1/

6 Community Resources

Resources that support of ***organizational*** and community building activities:

Tool	Location
Wiki for community collaboration and planning	https://nationaldataservice.atlassian.net/wiki/
Trello board to track high level organizational tasks	https://trello.com/b/oQRM22mi/nds-management https://trello.com/b/ROFp7iFi/nds-tac
Slack team for distributed group chat	https://nationaldataservice.slack.com
YouTube channel for demo/informational videos	https://www.youtube.com/channel/UCWuPo7LDCzsqF3RaKzIQmHw

Resources that support ***development*** efforts:

Tool	Location
GitHub repository for code sharing	https://github.com/nds-org
Confluence, a wiki, for planning, discussion, meeting notes, and other documentation associated with the development of the NDS Labs and Share resources	https://opensource.ncsa.illinois.edu/confluence/display/NDS/
Jira for task, issue, bug tracking, and coordinating development sprints	https://opensource.ncsa.illinois.edu/jira/projects/NDS/
Hipchat a development centric team chatroom	https://hipchat.ncsa.illinois.edu/g86jA5TZg
Discussion list for the development team	ncsa-nds-dev@nataionaldataservice.org

User support tools include:

Tool	Location
Google group for questions and answers	https://groups.google.com/forum/#!forum/ndslabs
Gitter chatroom	https://gitter.im/nds-org/ndslabs