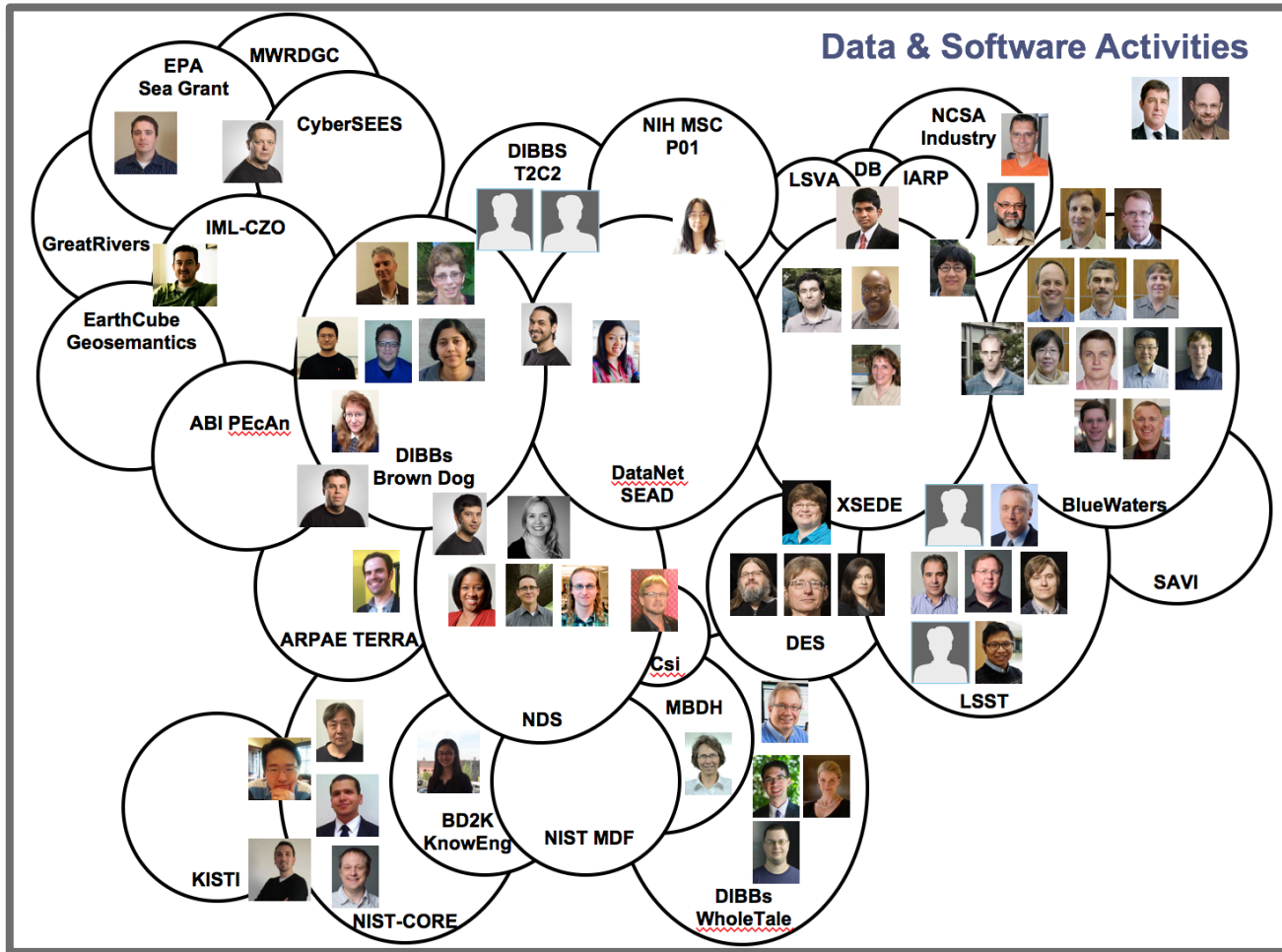


NATIONAL CENTER FOR SUPERCOMPUTING APPLICATIONS

Supporting Scientific Analysis – From Services to Gateways





Research Software Development & Management in Universities: Case Studies from Manchester's RSDS Group, Illinois' NCSA, and Notre Dame's CRC

Daniel S. Katz^{*}, Kenton McHenry^{*}, Caleb Reinking[†], and Robert Haines[‡]
^{*}National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, IL, USA
 Email: {dskatz, mchenry}@illinois.edu
[†]University of Notre Dame, Notre Dame, IN, USA
 Email: caleb.m.reinking.2@nd.edu
[‡]University of Manchester, Manchester, UK
 Email: robert.haines@manchester.ac.uk

Abstract—Modern research in the sciences, engineering, humanities, and other fields depends on software, and specifically, research software. Much of this research software is developed in universities, by faculty, postdocs, students, and staff. In this paper, we focus on the role of university staff. We examine three different, independently-developed models under which these staff are organized and perform their work, and comparatively analyze these models and their consequences on the staff and on the software, considering how the different models support software engineering practices and processes. This information can be used by software engineering researchers to understand the practices of such organizations and by universities who want to set up similar organizations and to better produce and maintain research software.

1. INTRODUCTION

Modern research in the sciences, engineering, humanities, and other fields depends on software, and specifically, research software. The NSF made 18,592 awards totaling \$9.6 billion to projects that mentioned “software” in their abstracts between 1995 and 2016 [1]. An examination of 40 papers in Nature from January to March 2016 showed that 32 explicitly mentioned software, with each paper mentioning an average of 6.5 software tools, almost all of which were research software [2]. Two surveys, of academics and Russell Group Universities in the UK [3] and members of the National Postdoctoral Association in the US [4] found that about 2/3 of respondents said they couldn’t do their research without software, while about 1/3 said they could, but it would be much more difficult, and only a few percent said it would make no difference.

Research software is most often produced by researchers themselves, often within academia, by faculty, staff, postdocs, and students. While the academic environment and culture have developed over hundreds of years, software is much more recent. Many software projects are also developed and maintained over a longer time period than the academic tenure of any postdoc or student, and sometimes longer than staff and faculty as well. And until fairly recently, most software was developed in an ad hoc manner, without frameworks and tools (such as GitHub, science gateways, developer environments,



In 2012, a small group in the UK attending the Software Sustainability Institute (SSI) Collaborations Workshop (<http://software.ac.uk/cw12>) came together to address the question: why is there no career for software developers in academia? [5] The group members then wrote about their ideas on the people they were talking about, calling them Research Software Engineers (RSEs), their potential career paths, how their work is funded, and the challenges they face. [6] The SSI’s policy challenge in 2013, which led to the formation of the UK Research Software Engineers’ Association (UK RSE), was to increase the number of members of the UK RSE from 1,275 in 2013 to 633 in August 2016 to 1,275 in 2017. UK RSEs typically fit one of two models: either embedded in research groups or as a central organization in their institutions to support various research projects on a short-term basis.



Research Data Challenges

- Storage
- Everything else!!!
 - The bytes are not enough on their own
00110100 00110010
 - Metadata, curation tools, indexes, storage abstraction, replication, data transfer, authentication, access control, transformation, analysis tools, ...

→ *Software*

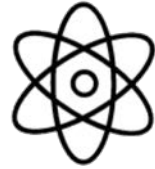
MarketMaker



CILogon

Evidence Visualizer

Parsl



4CEED



Omi



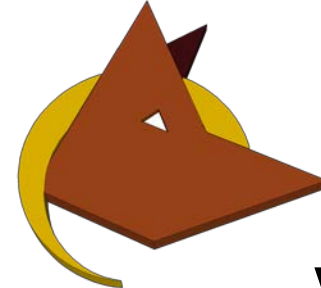
YesWorkflow x

MyProxy



WIX

Eclipse PTP



XRAS

NEST



Versus



Virtual Director



KNOWeng



3D Utilities

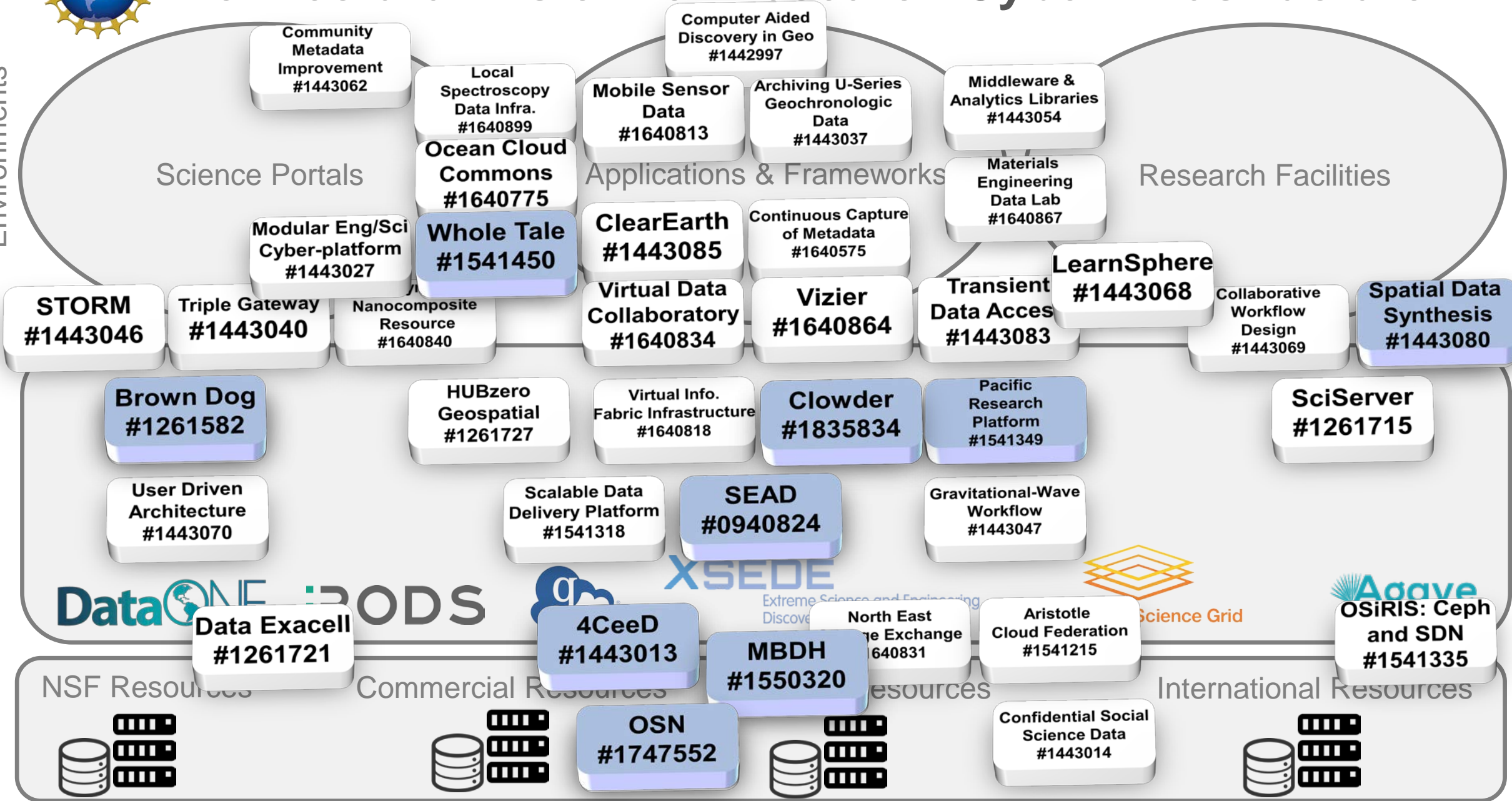


Architectural Vision for Research Cyberinfrastructure

Discipline Specific Environments

Integrative Services

Resources





Architectural Vision for Research Cyberinfrastructure

Discipline Specific Environments

Integrative Services

Resources

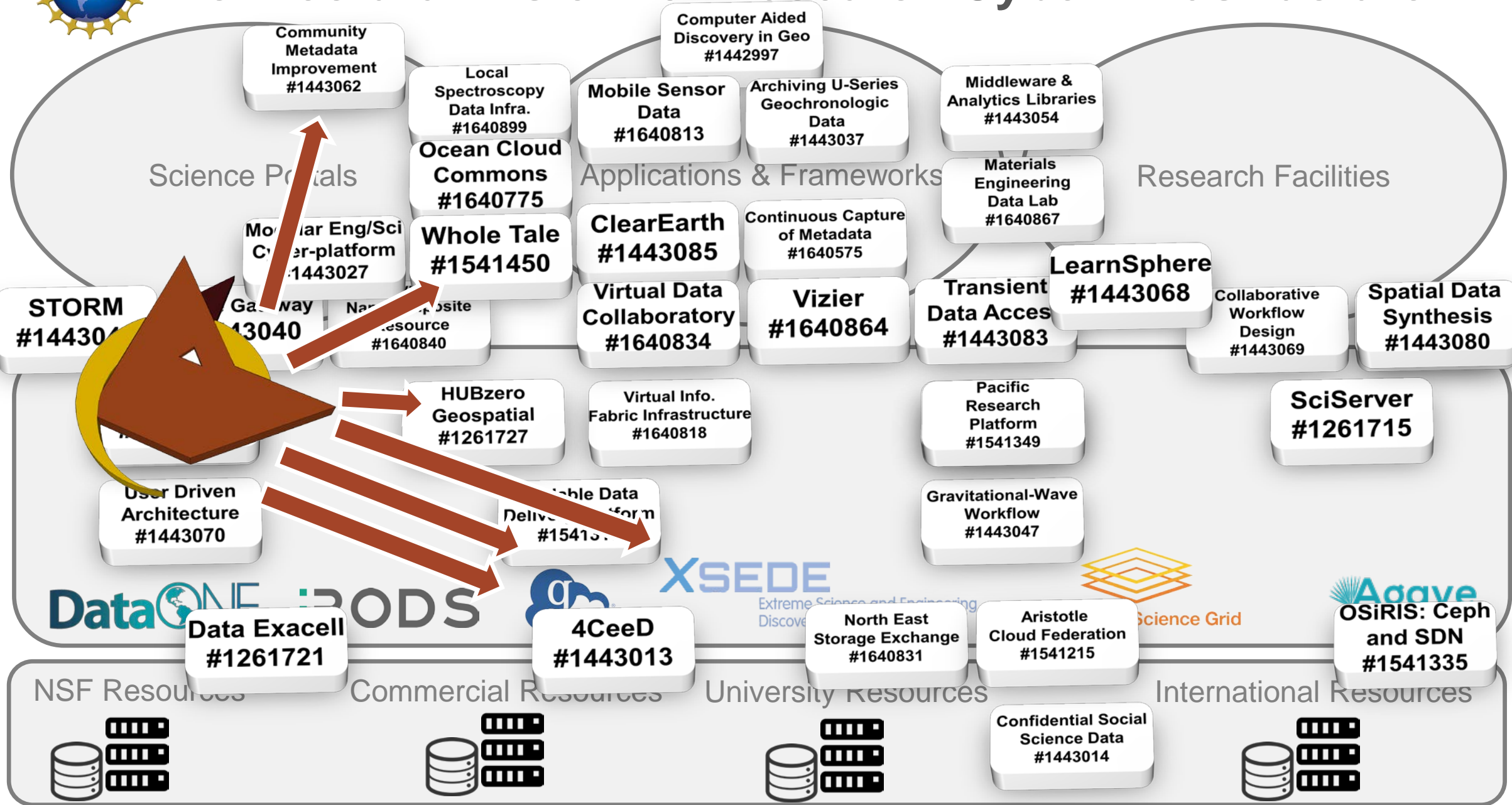
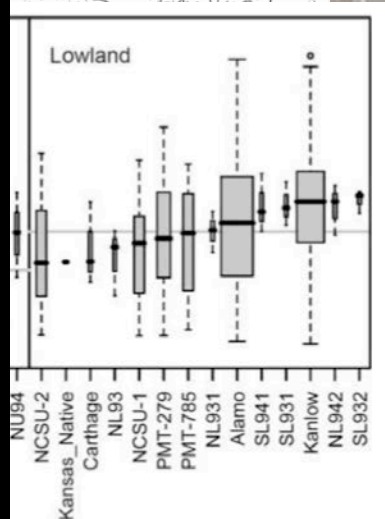
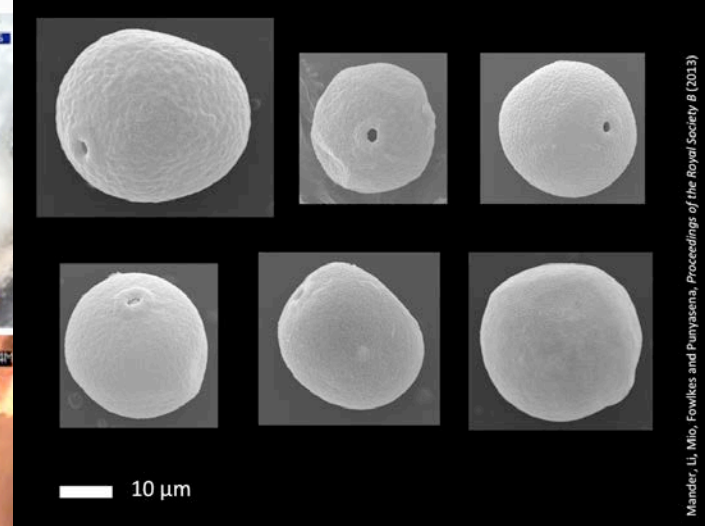
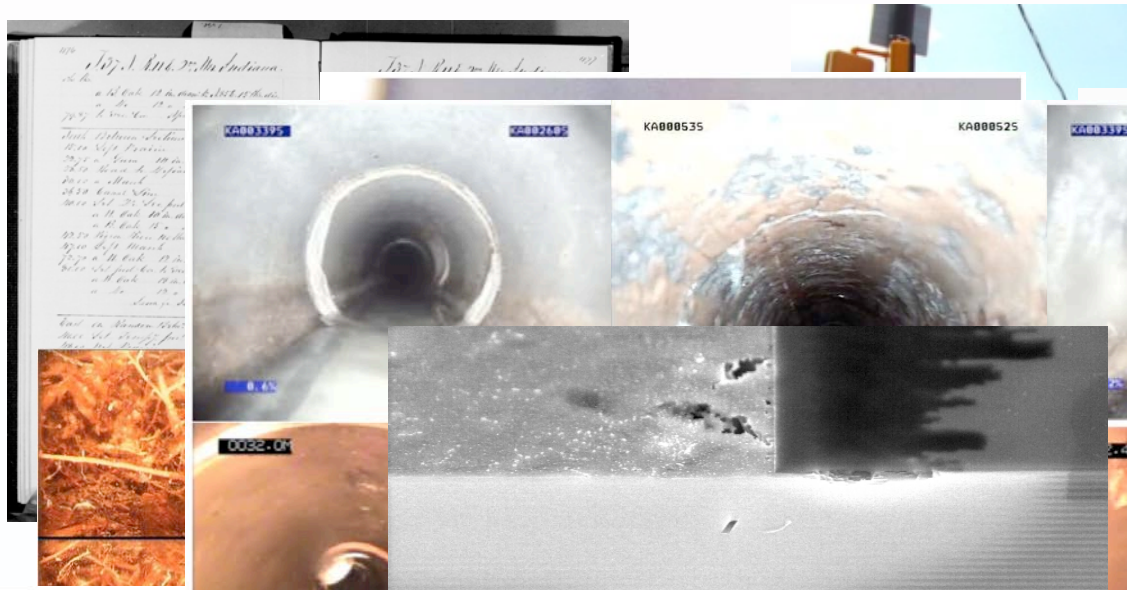


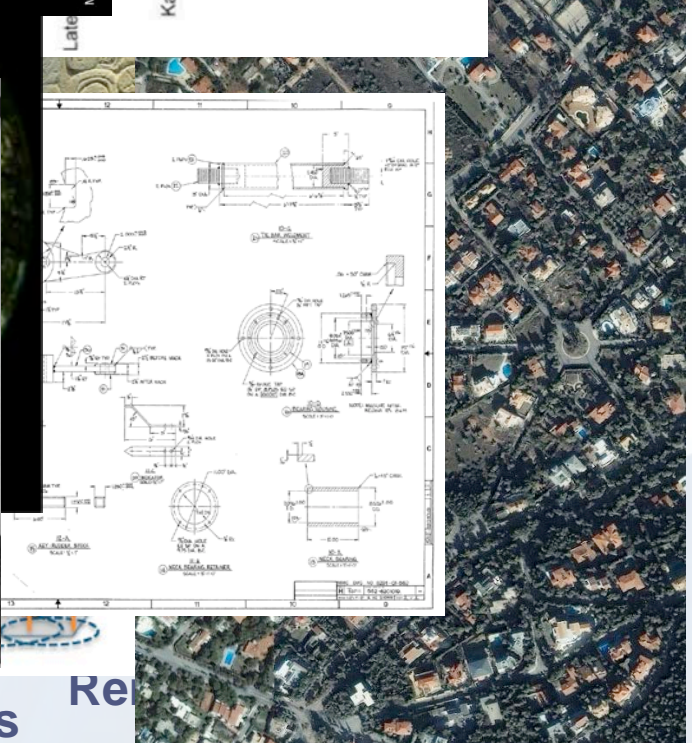
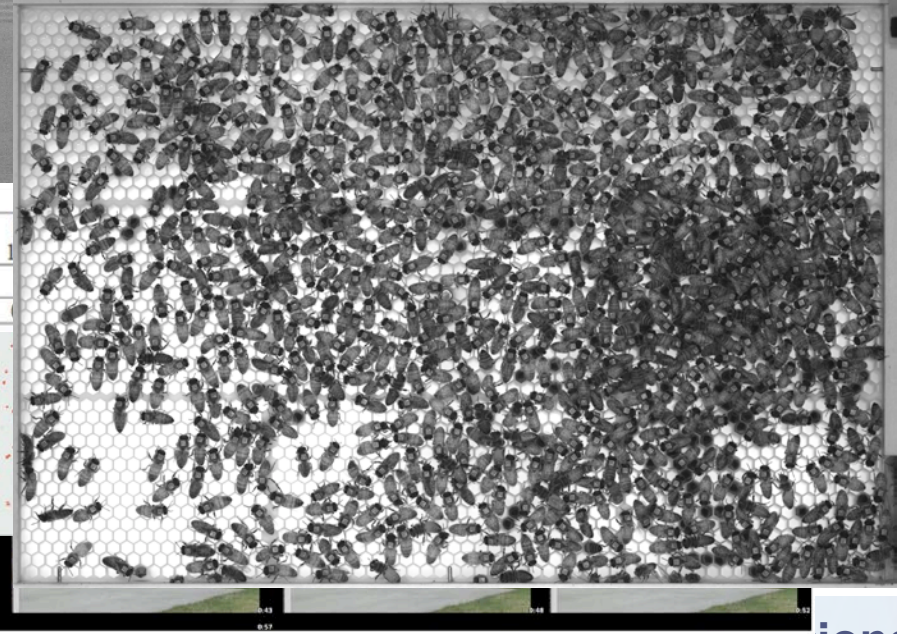
Image Stitching



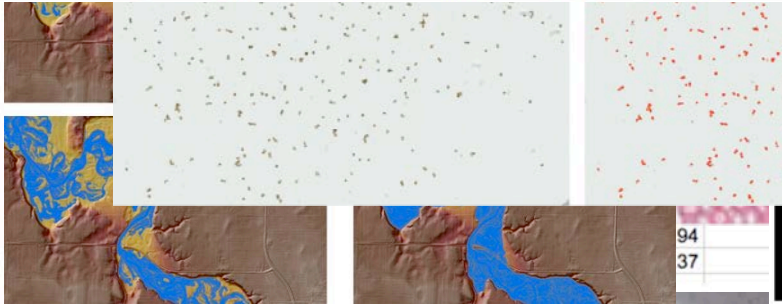
Indicators	N_0	
$t\ ha^{-1}$	15.8	
%	100	127
kg	1.66	2.26*

Annual biomass						
$t\ ha^{-1}$	27.0	28.5	29.7	5.31	10.5	10.7
%	100	105.6	110.1	18.68	100	102.3
Biomass weight per plant						
kg	2.05	2.18	2.25	0.396	0.79	0.81

10.0kV 7.4mm x9.00k



Co



3D Reconstruction

Event Detection

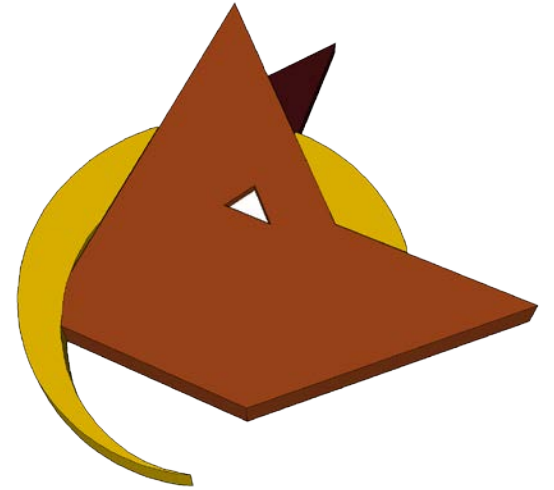
Simulations

Re

19



Conversion



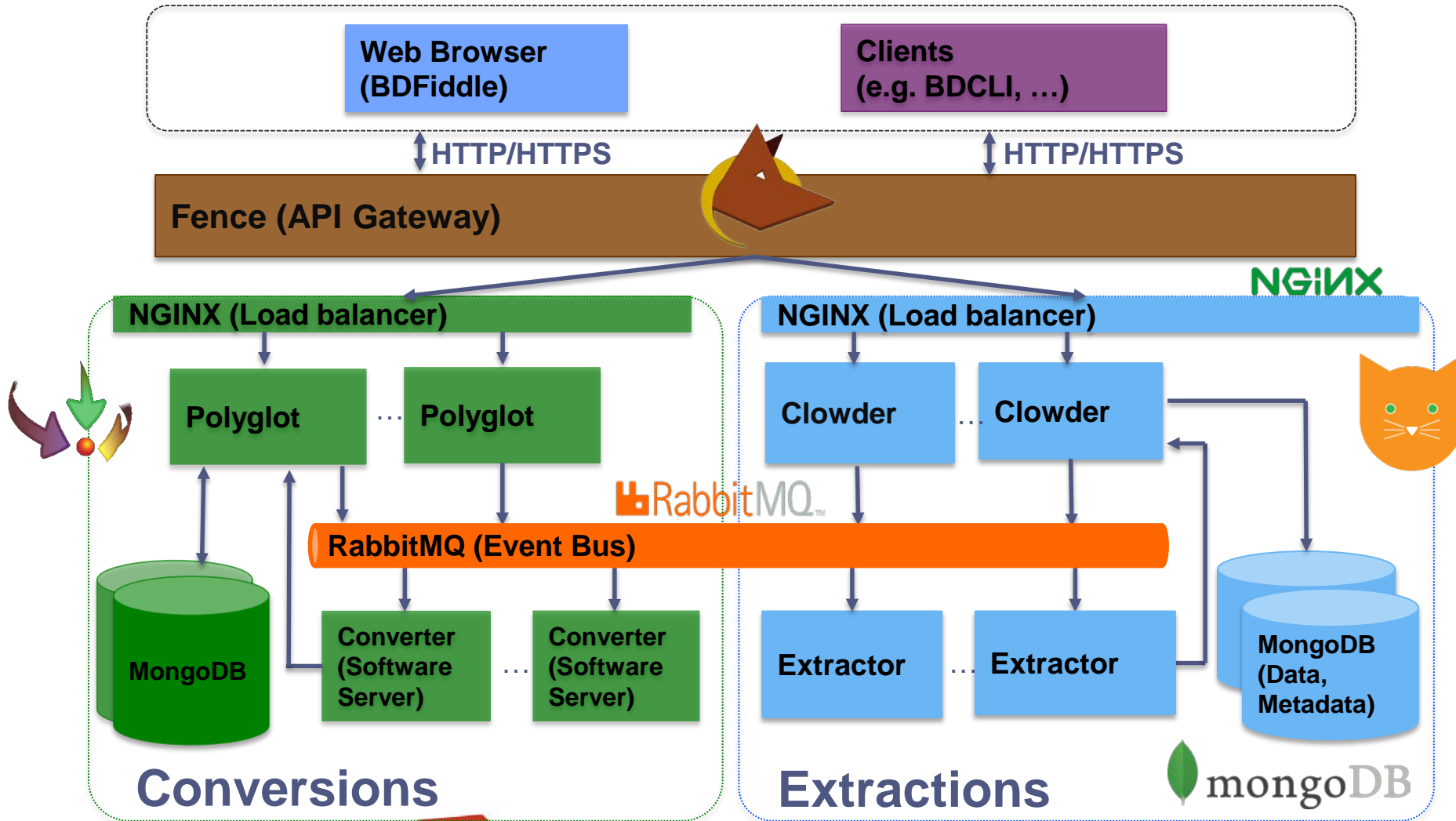
Extraction

```
{
  "extractor_id": "nca.image.exif",
  "Image": "558c3d84e4b00c3a039d5ac5",
  "Format": "JPEG (Joint Photographic Experts Group JFIF format)",
  "Class": "DirectClass",
  "Geometry": "2592x1936+0+0",
  "Resolution": "72x72",
  "Print size": "36x26.8889",
  "Units": "PixelsPerInch",
  "Type": "TrueColor",
  "Endianness": "Undefined",
  "Colorspace": "sRGB",
  "Depth": "8-bit",
  "Channel depth": {
    "red": "8-bit",
    "green": "8-bit",
    "blue": "8-bit"
  },
  "Channel statistics": {
    "Red": {
      "min": "0 (0)",
```

```
{
  "extractor_id": "nca.image.ocr",
  "ocr_simple": [
    "EB BROWSER MOSAIC THE FIRST POPULAR BROWSER FOR THE WORLD WIDE BY MARC ANDREESSEN BINA THE NATIONAL CENTER COMPUTING APPLICATIONS 1993 RELEASE TO THE PUBLIC INTERNET USERS EASY ACCESS TO SOURCES OF INFORMATION win HAVE TRANSFORMED THE INFORMATION UNIVERSITY OF "
  ],
  "Human Preference Extractor": {
    "Definitions": {
      "Human Preference": "A Computer Vision model that uses the spectral data of an image to get a human preference value ranging from 1 to 5.",
      "Green Index": "The green index is the estimated percentage of green pixels w"
    },
    "Human Preference": "4",
    "Green Index": "53.8"
  },
  "tags": [
    "Human Face Automatically Detected",
    "Person Automatically Detected",
    "Human Eyes Automatically Detected"
  ],
  "files": [
    "ts-dev.nca.illinois.edu:9000/files/558c3d84e4b00c3a039d5ac5"
  ],
  "datasets": [
    "ts-dev.nca.illinois.edu:9000/datasets/558c3dd6e4b00c3a039d5b77"
  ]
}
```

```
{
  "id": "558c3d84e4b00c3a039d5ac5",
  "filename": "IMG_0997.JPG",
  "tags": [
    "Human Face Automatically Detected",
    "Person Automatically Detected",
    "Human Eyes Automatically Detected"
  ]
}
```


DIBBs Brown Dog



Tools – Convertors & Extractors

PEcAn#ED_convert.R

```
#!/usr/bin/Rscript
#PEcAn
#data
#pecan.zip
#ed.zip

.libPaths("/home/polyglot/R/library")
sink(stdout(),type="message")

# global variables
overwrite <- TRUE
verbose <- TRUE

# get command line arguments
args <- commandArgs(trailingOnly = TRUE)

usage <- function(msg) {
  print(msg)
  print(paste0("Usage:      ", args[0], "
  print(paste0("Example1: ", args[0], "
...

```

wordcount.py

```
import subprocess

def wordcount(input_file):
    result = subprocess.check_output(['wc', input_file], stderr=subprocess.STDOUT)
    (lines, words, characters, _) = result.split()
    metadata = {
        'lines': lines,
        'words': words,
        'characters': characters
    }

    result = {
        'metadata': metadata
    }

    return result

```

Welcome to the Brown Dog Tools Catalog, a community resource for publishing data transformation tools and making them available for use by the scientific community. To add a new tool click "Contribute".

Tools:

Filter by tool types: Filter by tool levels:

Displaying 1 to 24 of totally 42 available:

The Predictive Ecosystem Analyzer

Convert data sources to formats required by community built ecological models

Tool type: Converter

Tool level: **Level 4**

Deployments: bd-api bda-api-dev

Submitter: mchenry

Last modified : Nov 23, 2016

GI Detector

Identify GI types in satellite images

Tool type: Extractor

Tool level: **Level 4**

Deployments: bd-api-dev

Submitter: mchenry

Last modified : Nov 22, 2016

FFmpeg

Convert between various video formats

Tool type: Converter

Tool level: **Level 4**

Submitter: mchenry

Last modified : Nov 20, 2016

Daffodil

Convert the contents of arbitrary files into an XML representation

Tool type: Converter

Tool level: **Level 4**

Deployments: bd-api bda-api-dev

Submitter: mchenry

Last modified : Nov 20, 2016

Ghostscript

Convert between various document formats

Tool type: Converter

Tool level: **Level 4**

Deployments: bd-api bda-api-dev

Submitter: mchenry

Last modified : Nov 20, 2016

OpenJPEG

Convert JPEG2000 images

Tool type: Converter

Tool level: **Level 4**

Deployments: bd-api-dev

Submitter: mchenry

Last modified : Nov 20, 2016

Kabeia

https://bd-api.ncsa.illinois.edu

 swagger

Authorize

Brown Dog API

Documentation for the [Brown Dog API](#)

Created by Brown Dog Support

See more at <https://opensource.ncsa.illinois.edu/confluence/questions/topics/41156618/brown-dog>

[Contact the developer](#)

[University of Illinois/NCSA Open Source License](#)

Authorization

Show/Hide | List Operations | Expand Operations

POST	/keys	Create a new api key
POST	/keys/{key}/tokens	Create a new access token from api key
DELETE	/keys/{key}	Delete api key and all related access tokens
DELETE	/tokens/{token}	Delete token
GET	/tokens/{token}	

Conversions

Show/Hide | List Operat

GET	/dap/outputs	List all output formats that can be reached
-----	--------------	---

powered by 
BROWN DOG

https://bd-api.ncsa.illinois.edu

DELETE	/keys/{key}	Create a new access token from api key
DELETE	/tokens/{token}	Delete token

GET	/tokens/{token}	Get info about token
-----	-----------------	----------------------

Conversions

Show/Hide | List Operations | Expand Operations

GET	/dap/outputs	List all output formats that can be reached
-----	--------------	---

GET	/dap/inputs	List all input formats that can be accepted. Limit the number of hops allowed by setting chain parameter, e.g. ?chain=2
-----	-------------	---

GET	/dap/inputs/{input_format}	List all output formats that can reach the specified input format
-----	----------------------------	---

GET	/dap/convert	List all output formats that can be reached
-----	--------------	---

GET	/dap/convert/{output_format}	List all input formats that can reach the specified output format
-----	------------------------------	---

POST	/dap/convert/{output_format}	Convert the uploaded file to the requested output format
------	------------------------------	--

GET	/dap/convert/{output_format}/{file_url}	Convert the specified file to the requested output format
-----	---	---

GET	/dap/path/{output_format}/{input_format}	Return the conversion path that would be used for this conversion
-----	--	---

GET	/dap/software	List all available conversion software
-----	---------------	--

GET	/dap/software/{software}	List all output formats available for this software
-----	--------------------------	---

GET	/dap/software/{software}/{output_format}	List all input formats available for this software
-----	--	--

POST	/dap/software/{software}/{output_format}	Convert the uploaded file to the requested output format using the specified software
------	--	---

GET	/dap/software/{software}/{output_format}/{file_url}	Convert the specified file to the requested output format using the specified software
-----	---	--





A Science Driven Data Transformation Service

BD Fiddle

BD Catalog

Clients

Sign Up!

63,029,921 files/datasets transformed



```
else:  
    print "HTTP status code: " + response.status_code
```

1 2 3
Select File Select Transformation Copy Code Snippet Download Code



```
# Make sure that that path to the input file is correct.  
library(BrownDog)  
  
output_file <- convert_file("https://bd-api.ncsa.illinois.edu", "IMG0008.PCD", "png", "./",  
"e941ba85-8020-4115-9512-09ebbf7ab156")
```

Open Jupyter Notebook Download Jupyter Notebook Copy To Clipboard Download Code



```
%% Make sure that that path to the input file is correct.  
  
bd_funcs = bd_client;  
  
output_file = bd_funcs.convert('https://bd-api.ncsa.illinois.edu', 'IMG0008.PCD', 'png', './  
/converted_file.png', 'e941ba85-8020-4115-9512-09ebbf7ab156');
```

Copy To Clipboard Download Code



```
The bd.js library requires URLs.
```

Copy To Clipboard Download Code

File (<5M): URL:
IMG_0997.jpg 1 2 3
Select File Select Transformation Token (Logout)
248f2ca6-be58-446e-800b-4ef817c5e4c7
Copy Code Snippet
Get Transformations

2) Select Transformation:

Automatic **Manual** **Convert:**

- Clowder
- GDAL
- GeoExpress
- ImageMagick
- OpenJPEG

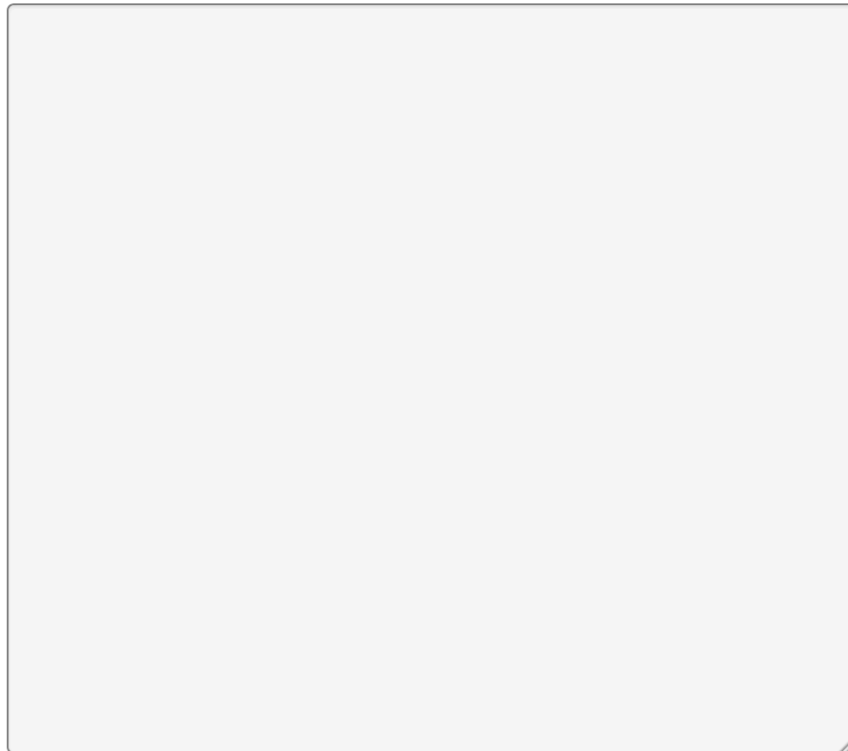
Extract:

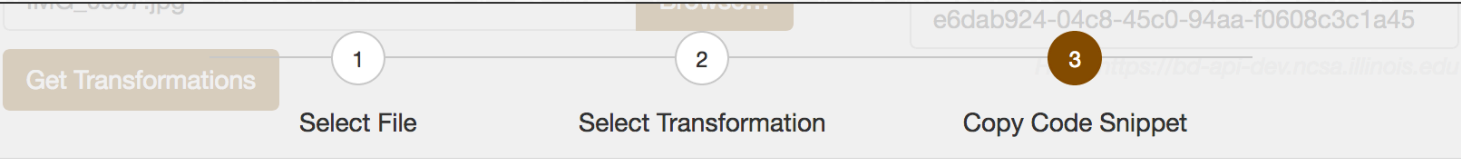
- gi_detector
- ncsa.cv.caltech101
- ncsa.cv.closeups
- ncsa.cv.eyes
- ncsa.cv.faces
- ncsa.cv.meangrey
- ncsa.cv.profiles
- ncsa.image.metadata
- ncsa.image.ocr
- ncsa.image.preview

Email Output Big Data
 Generate workflow

Submit

Output:





2) Select Transformation:

Automatic **Convert:**

Manual

- ascii
- azw3
- bil.zip
- bmp
- clim
- csv
- dalec
- dib
- doc
- docx

Extract:

image/jpeg

Email Output Big Data

Generate workflow

Submit

Output:

```
"Human Preference": {
  "Human Preference": 4,
  "Human Preference Definition": "An index [1,5]
from a GI model that rates human preference for Green I
nfrastructure"
},
"Green Index": {
  "Green Index Definition": "The green index is t
he estimated percentage of green pixels within an image
.",
  "Green Index": "53.8"
}
},
{
"@context": [
  "https://clowder.ncsa.illinois.edu/contexts/metad
ata.jsonld",
  {
    "eyes": "http://clowder.ncsa.illinois.edu/ncsa.
cv.eyes#eyes"
```

3) Code Snippets:

```
else:  
    print "HTTP status code: " + response.status_code
```

1 2 3

Select File Select Transformation Copy Code Snippet Download Code



```
# Make sure that that path to the input file is correct.  
library(BrownDog)  
  
metadata <- extract_file("https://bd-api-dev.ncsa.illinois.edu", "IMG_0997.jpg", "e6dab924-04c8-45c0-94aa-f0608c3c1a45")
```

Open Jupyter Notebook Download Jupyter Notebook Copy To Clipboard Download Code



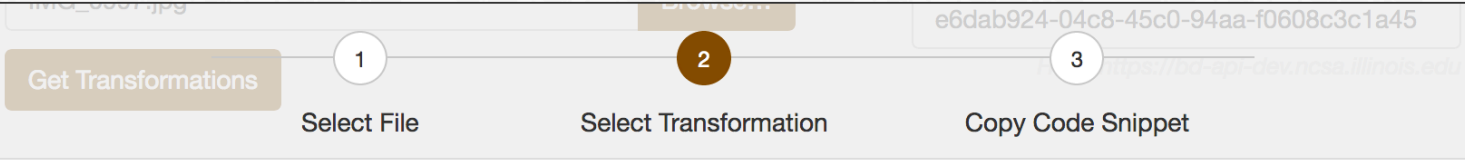
```
%% Make sure that that path to the input file is correct.  
  
bd_funcs = bd_client;  
  
metadata = bd_funcs.extract('https://bd-api-dev.ncsa.illinois.edu', 'IMG_0997.jpg', 'e6dab924-04c8-45c0-94aa-f0608c3c1a45');
```

Copy To Clipboard Download Code



```
The bd.js library requires URLs.
```

Copy To Clipboard Download Code



2) Select Transformation:

Automatic **Convert:**

Manual

- Clowder
- GDAL
- GeoExpress
- ImageMagick
- OpenJPEG

Extract:

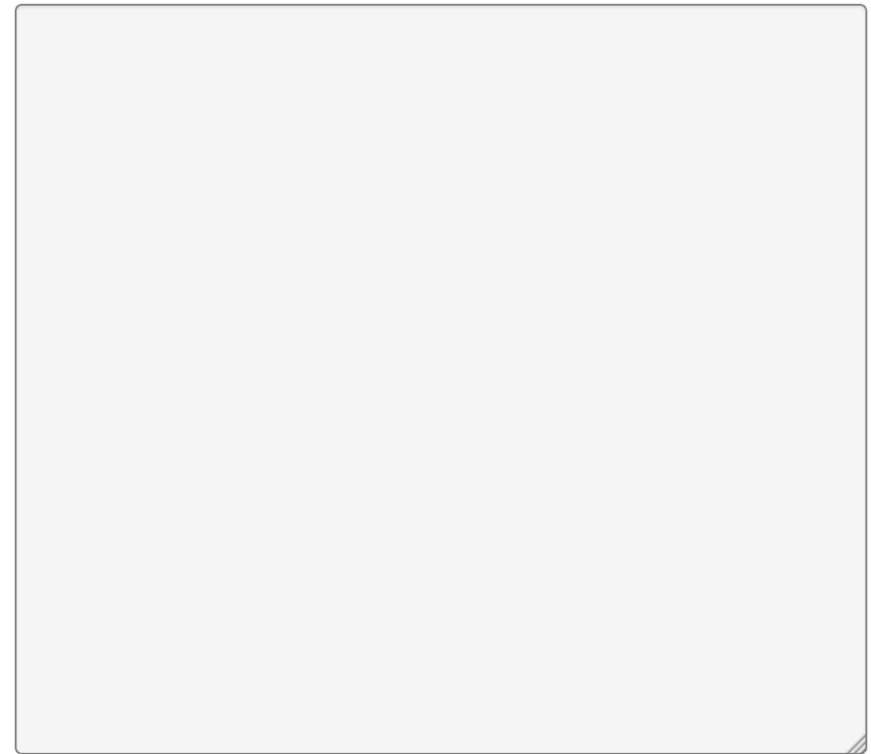
- gi_detector
- ncsa.cv.caltech101
- ncsa.cv.closeups
- ncsa.cv.eyes
- ncsa.cv.faces
- ncsa.cv.meangrey
- ncsa.cv.profiles
- ncsa.humanpref**
- ncsa.image.metadata
- ncsa.image.ocr

Email Output Big Data

Generate workflow

Submit

Output:



3) Code Snippets:

PECAn Parameter Selection

pecan.ncsa.illinois.edu/pecan/03-inputs.php

Selected Site

Set parameters for the run.

PFT*
 populus
 temperate.coniferous
 temperate.deciduous

Start Date*
 2004/01/01

End Date*
 2004/12/31

Sipnet.climna*
 Use Ameriflux

Email

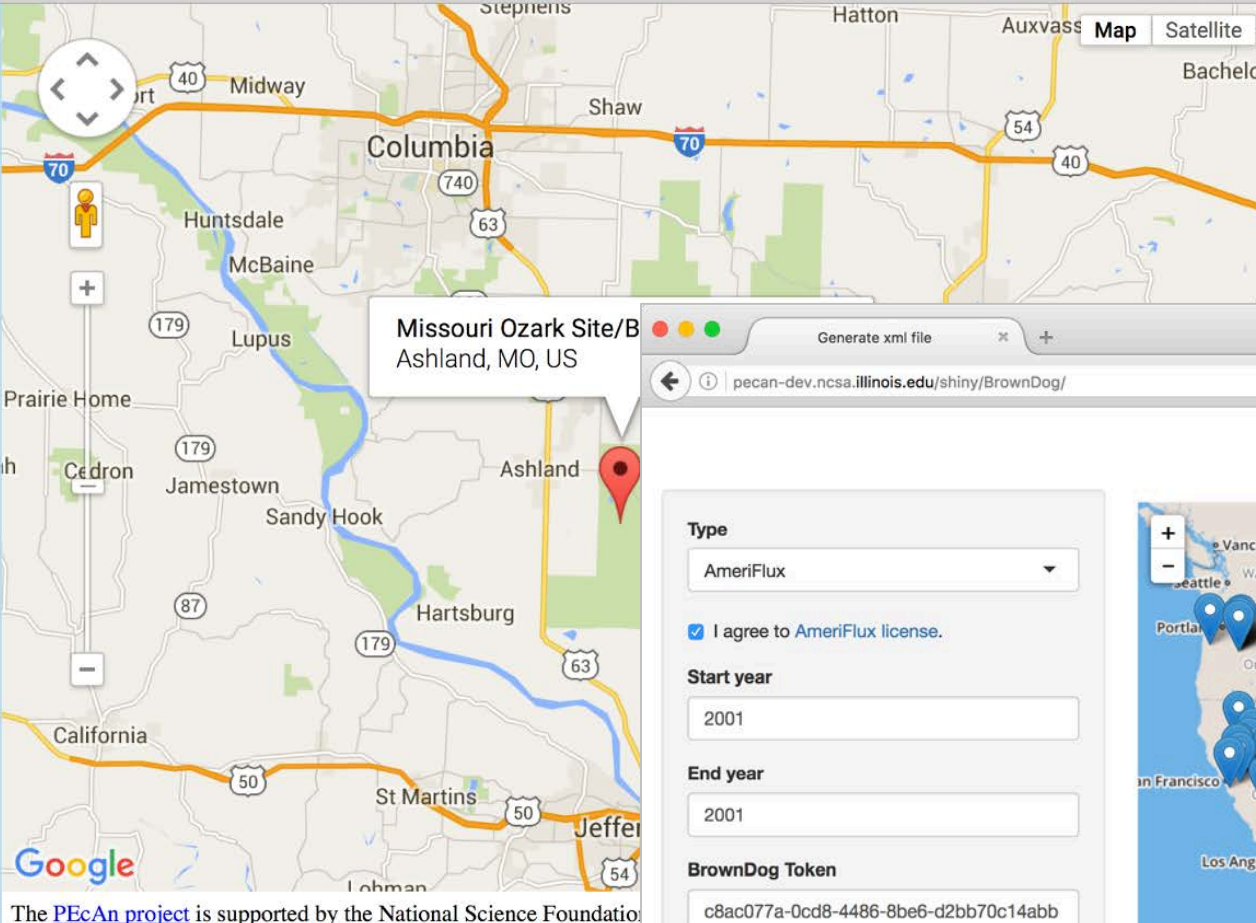
Use [BrownDog](#)

Edit pecan.xml

Edit model config

Advanced setup

The [PECAn project](#) is supported by the National Science Foundation



Missouri Ozark Site/B
Ashland, MO, US

Generate xml file

pecan-dev.ncsa.illinois.edu/shiny/BrownDog/

Type
 AmeriFlux

I agree to AmeriFlux license.

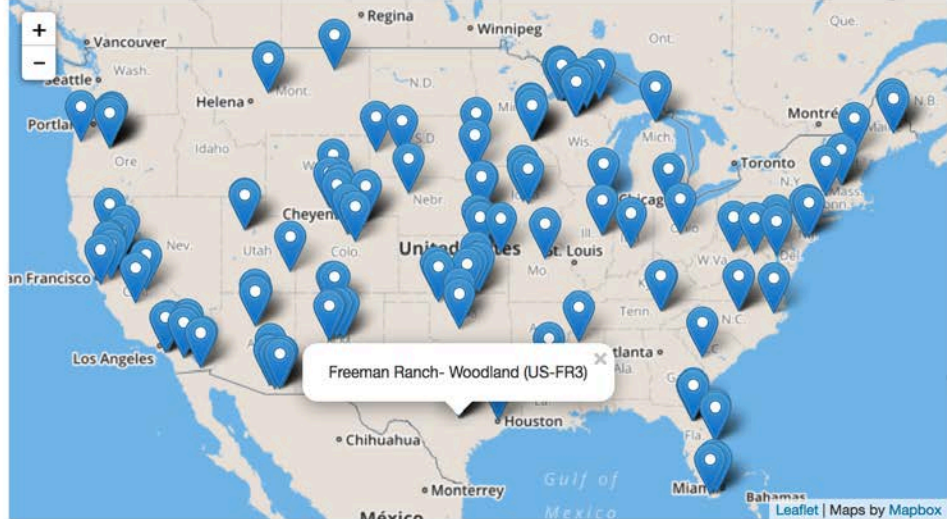
Start year
 2001

End year
 2001

BrownDog Token
 c8ac077a-0cd8-4486-8be6-d2bb70c14abb

Model
 ed.zip

ameriflux.zip
 clim
 dalec
 ed.zip
 linkages
 pecan.nc
 pecan.zip



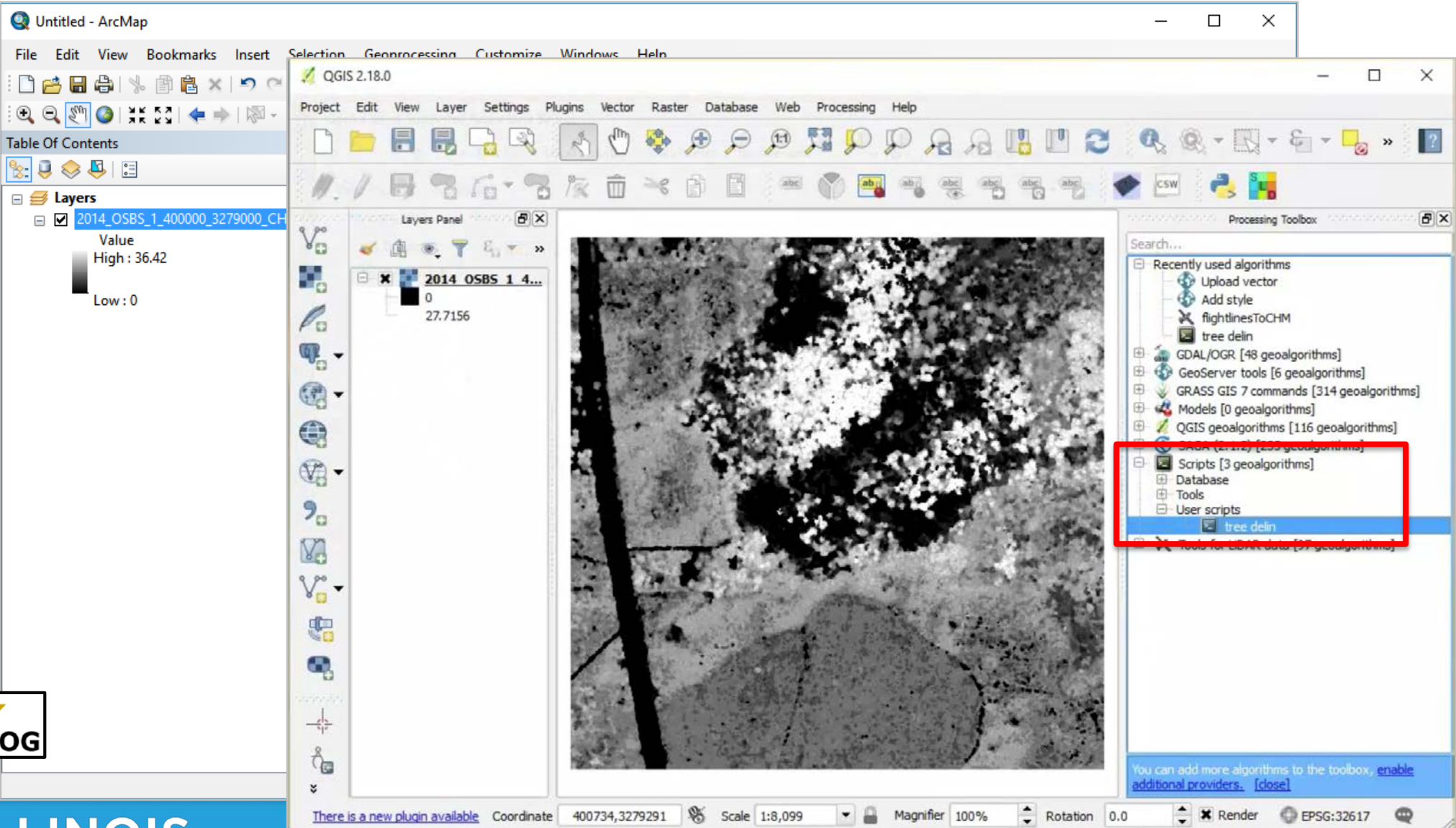
Freeman Ranch- Woodland (US-FR3)

```
<input>
<type>AmeriFlux</type>
<site>Freeman Ranch- Woodland (US-FR3)</site>
<lat>29.94</lat>
<lon>-97.99</lon>
<start_date>2001-01-01 00:00:00</start_date>
<end_date>2001-12-31 23:59:59</end_date>
</input>
```

Download XML Download Data



Geospatial Software



General Software

The screenshot displays a Microsoft Excel spreadsheet with a table of route data. A red box highlights the data from row 1 to row 13, columns A to E. To the right of the spreadsheet is a 'Brown Dog' service panel titled 'Green Route Index'. The panel contains instructions for using the service, a text input field for an access token, a 'Submit Selection' button, and a list of generated metadata links.

Route	Lat	Long	Green Index	Orientation
1	40.11123	-88.17377	30.8	0
1	40.11123	-88.17377	37.7	90
1	40.11123	-88.17377	38.9	180
1	40.11123	-88.17377	27.6	270
1	40.11202	-88.17488	32.1	270
1	40.11202	-88.17488	20.9	180
1	40.11202	-88.17488	32.2	0
1	40.11202	-88.17488	47.1	90
1	40.11289	-88.17517	31.6	0
1	40.11289	-88.17517	16.9	90
1	40.11289	-88.17517	19.9	270
1	40.11289	-88.17517	24.1	180

Brown Dog

Green Route Index

1. Provide an Access Token:
2. Select cells in the sheet as a 2 column list of latitude, longitude pairs along a path.
3. Submit to Brown Dog service. A new sheet will be added to the workbook with the results once they are ready.

Submitted 6 rows and 2 columns
Adding metadata to sheet

[Original File Metadata](#)

- [Derived File 0 metadata](#)
- [Derived File 1 metadata](#)
- [Derived File 2 metadata](#)
- [Derived File 3 metadata](#)
- [Derived File 4 metadata](#)
- [Derived File 5 metadata](#)
- [Derived File 6 metadata](#)
- [Derived File 7 metadata](#)
- [Derived File 8 metadata](#)
- [Derived File 9 metadata](#)
- [Derived File 10 metadata](#)
- [Derived File 11 metadata](#)



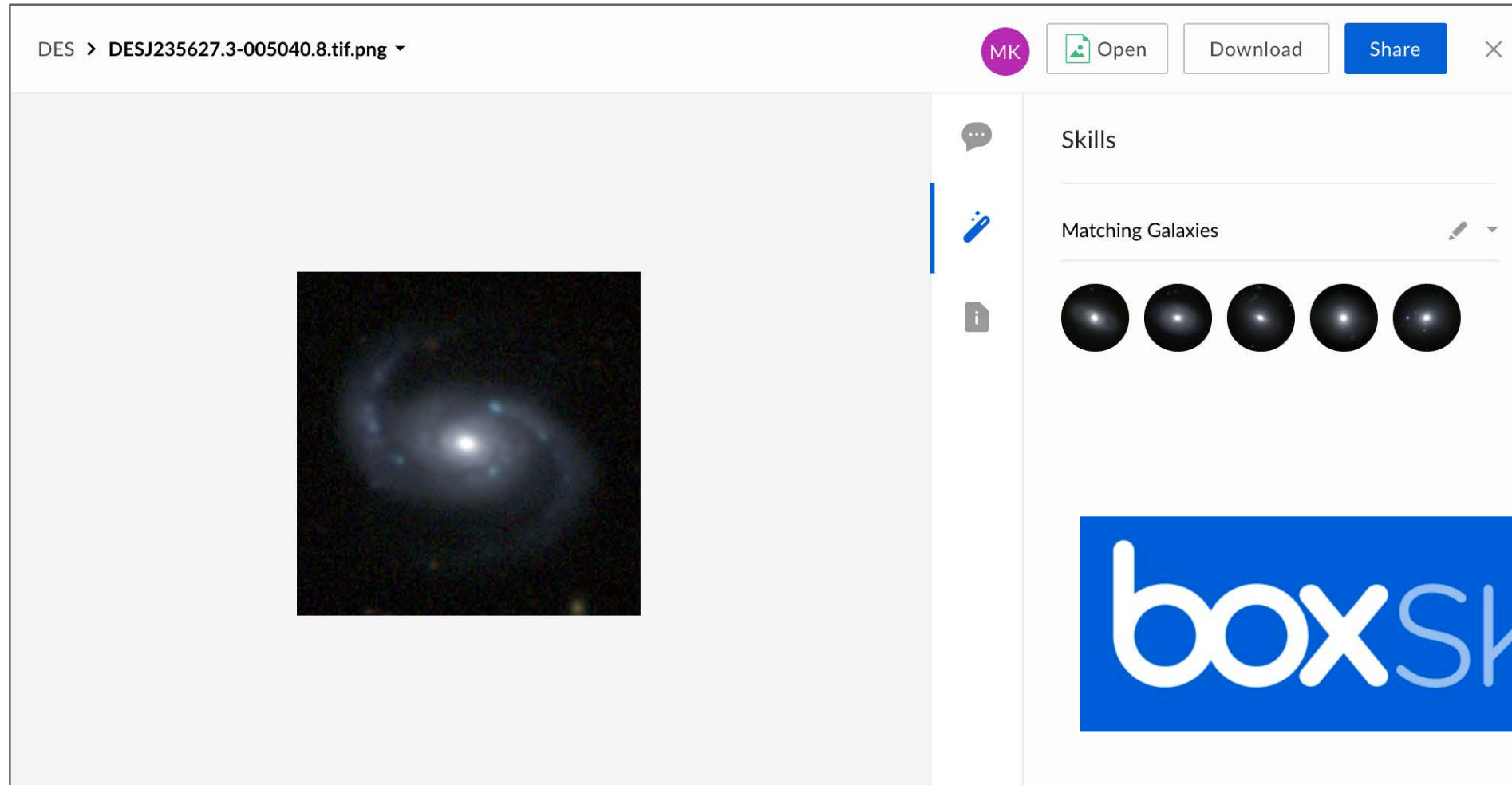
Box Integration

DES > DESJ235627.3-005040.8.tif.png

MK Open Download Share

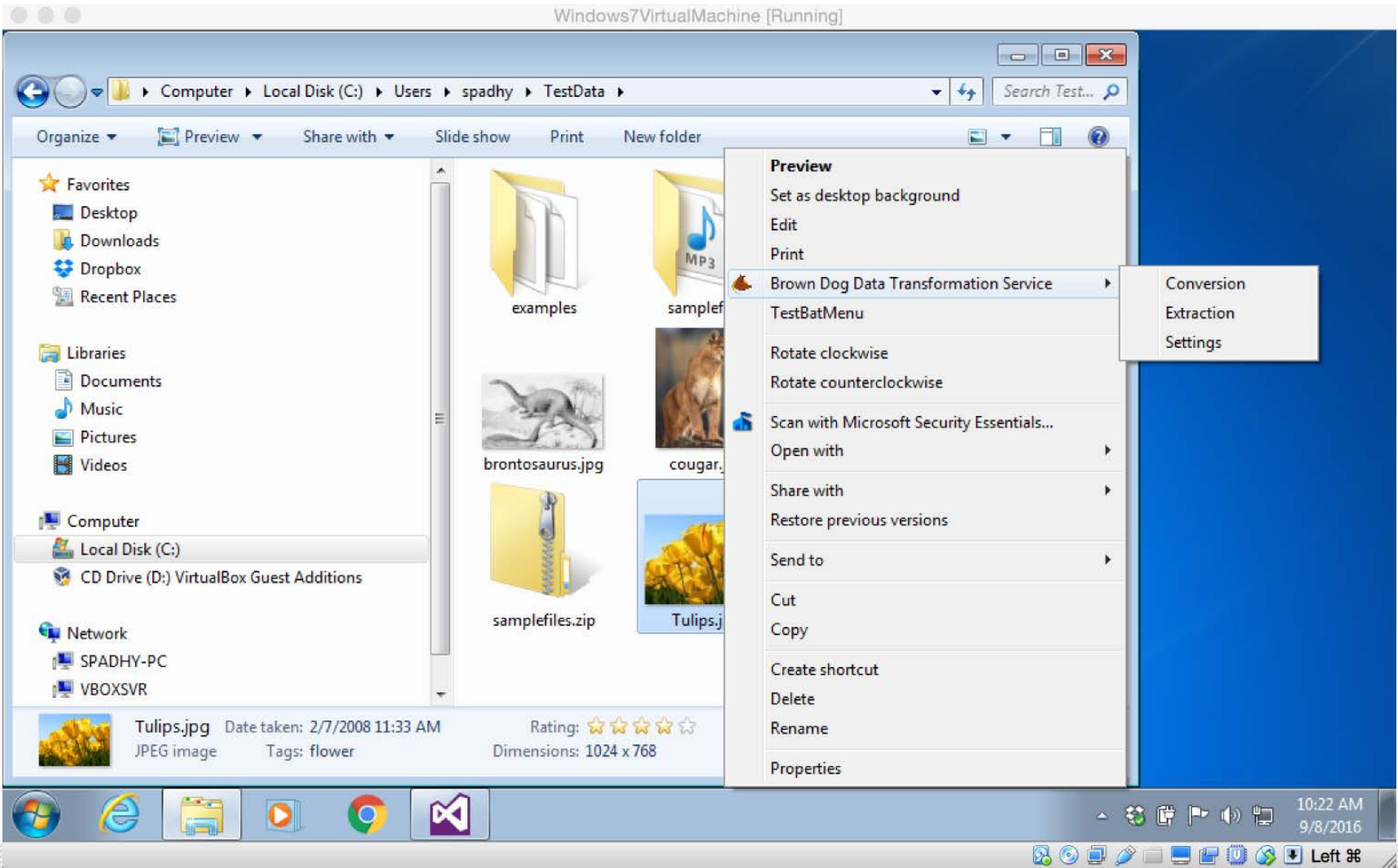
Skills

Matching Galaxies

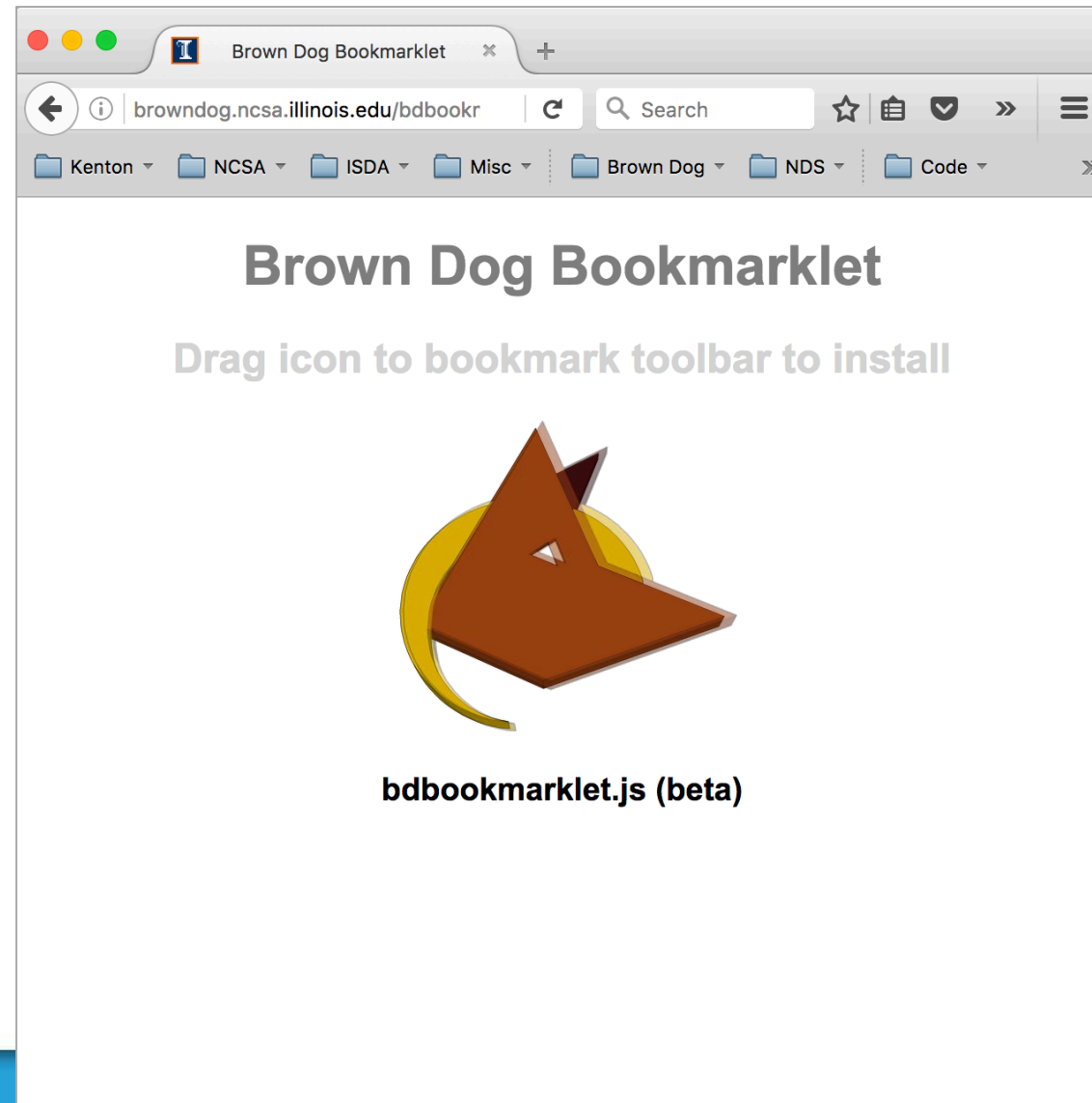


boxSKILLS

Operating Systems



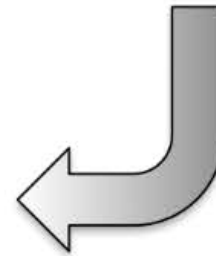
BD Bookmarklet



BD Bookmarklet

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
 Parent Directory		-	
 0005.jp2	2014-07-03 13:49	633K	
 0094.jp2	2014-07-03 15:15	7.6M	
 105_1.sid	2014-07-03 13:47	3.7M	
 aac aif aiff ascii asf avi azw3 bmp clim dalec dib doc docx ed.zip eps epsi epub			
 City Limits.7z	2014-10-01 12:43	763	
 Human_knownGene_chr22_1-51304566.bed	2015-04-14 13:57	260K	
 IMG_0997.jpg	2016-07-03 23:22	2.1M	
 US-Dk3-2001-2003.xml	2015-07-08 22:31	191	
 alice.txt	2014-12-27 23:06	141K	

Conversion menus added to links



BD Bookmarklet

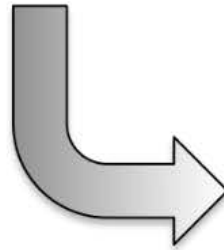
The screenshot shows a web browser window with the following elements:

- Address bar: `https://www.dropbox.com/s/w00so6d6tobq98n/Dietze_BrownDog_2014.odp`
- Browser tabs: "Dietze_BrownDog_2014.o...", "DAP", and others.
- Dropbox interface: Shows the file "Dietze_BrownDog_2014.odp" uploaded "3 days ago" with a size of "24.21 MB".
- Buttons: "Download" and "Save to my Dropbox".
- File format menu: A horizontal list of file formats including `ascii`, `azw3`, `bmp`, `dib`, `doc`, `docx`, `eps`, `epsi`, `epub`, `fb2`, `gif`, `html`, `htmlz`, `jp2`, `jpeg`, `jpg`, `lit`, `lrf`, `mobi`, `ods`, `odt`, `pdb`, `pdf`, `pdf12`, `pdf13`, `pdf14`, `pdfwr`, `pgm`, `pic`.

At the bottom of the browser window, a URL is visible: `kgm-d3.ncsa.illinois.edu:8184/convert/pdf/https://www.dropbox.com/s/w00so6d6tobq98n/Dietze_BrownDog_2014.odp?dl=1`

BD Bookmarklet


Press Ctrl+F to index and search data on page




Search Data ✕


Queryable Data Create Index 0 0 0

Search query Search

 **Image size**
2592 x 1936

 **Image size**
480 x 640

<http://brwonder.nps.gov/illinois/evangelical/brwonder.nps> **Link 1**

 Close

Lines of Code	273
Other files	Model Image Sample
Dependencies	numpy, argparse, glob, cv2, cPickle, random, h5py, skimage, sklearn, scipy
Difficulties	Install OpenCV (cv2)



```

exposure

and parse the arguments

", required=True,

"size",

help="Flag indicating whether or not to visualize each iteration")
ap.add_argument("-p", "--path",
                help="path to gi_detector")
args = vars(ap.parse_args())

path_1 = '/Users/ankitrai/Dropbox/ppao_VM/gi_detector/'
def pyramid(image, scale=1.5, minSize=(55,55)):
    # yield the original image
    yield image

    # keep looping over the pyramid
    while True:
        # compute the new dimensions of the image and resize it
        w = int(image.shape[1] / scale)
        image = imutils.resize(image, width=w)

        # if the resized image does not meet the supplied minimum
        # size, then stop constructing the pyramid
        if image.shape[0] < minSize[1] or image.shape[1] < minSize[0]:
            break

        # yield the next image in the pyramid
        yield image
def sliding_window(image, stepSize, windowSize):
    # slide a window across the image
    for y in xrange(0, image.shape[0], stepSize):
        for x in xrange(0, image.shape[1], stepSize):
            # yield the current window
            yield (x, y, image[y:y + windowSize[1], x:x + windowSize[0]])
def detect(image, winDim, winStep=4, pyramidScale=1.5, minPr

    # initialize the list of bounding boxes and associat

    # loop over the image pyramid
    pyramid_layers = pyramid(image, scale=1.5, minSize=(100,100))
    for layer in pyramid_layers:
        # determine the current scale of the pyramid
        scale = image.shape[0] / float(layer.shape[0])
        # loop over the sliding windows for the current pyramid layer
        for (x, y, window) in sliding_window(layer, winStep, winDim):
            (winH, winW) = window.shape[:2]
            if winH == winDim[1] and winW == winDim[0]:
                # extract HOG features from the current window and classify

```

```

# ... (code from another file) ...

```

Total Code from 2 Files

Lines of Code	47
Other files	None
Dependencies	bd, requests, os, glob, argparse, time, json, PIL
Difficulties	

```

import sys
import requests
import os
import argparse
import time
import json
from PIL import Image, ImageDraw
from glob import glob

# Construct the argument parser and parse the arguments
ap = argparse.ArgumentParser()
ap.add_argument("-t", "--token", required=True, help="API token")
ap.add_argument("-i", "--images", help="Path to image directory")
ap.add_argument("-b", "--bdapi", help="Brown Data API URL")
args = ap.parse_args()

# Parse the arguments
token = args["token"]
if args["images"] != None:
    image_path = args["images"]
else:
    image_path = "./images"
if args["bdapi"] != None:
    bdapi = args["bdapi"]
    if bdapi[-1] != "/":
        bdapi = bdapi + "/"
else:
    bdapi = "https://bd-api.ncsa.illinois.edu/"

# Loop through images in image directory
for image_path in glob.glob(image_path + "/*.jpg"):
    metadata = extract(bdapi, image_path, token)['metadata.jsonld']

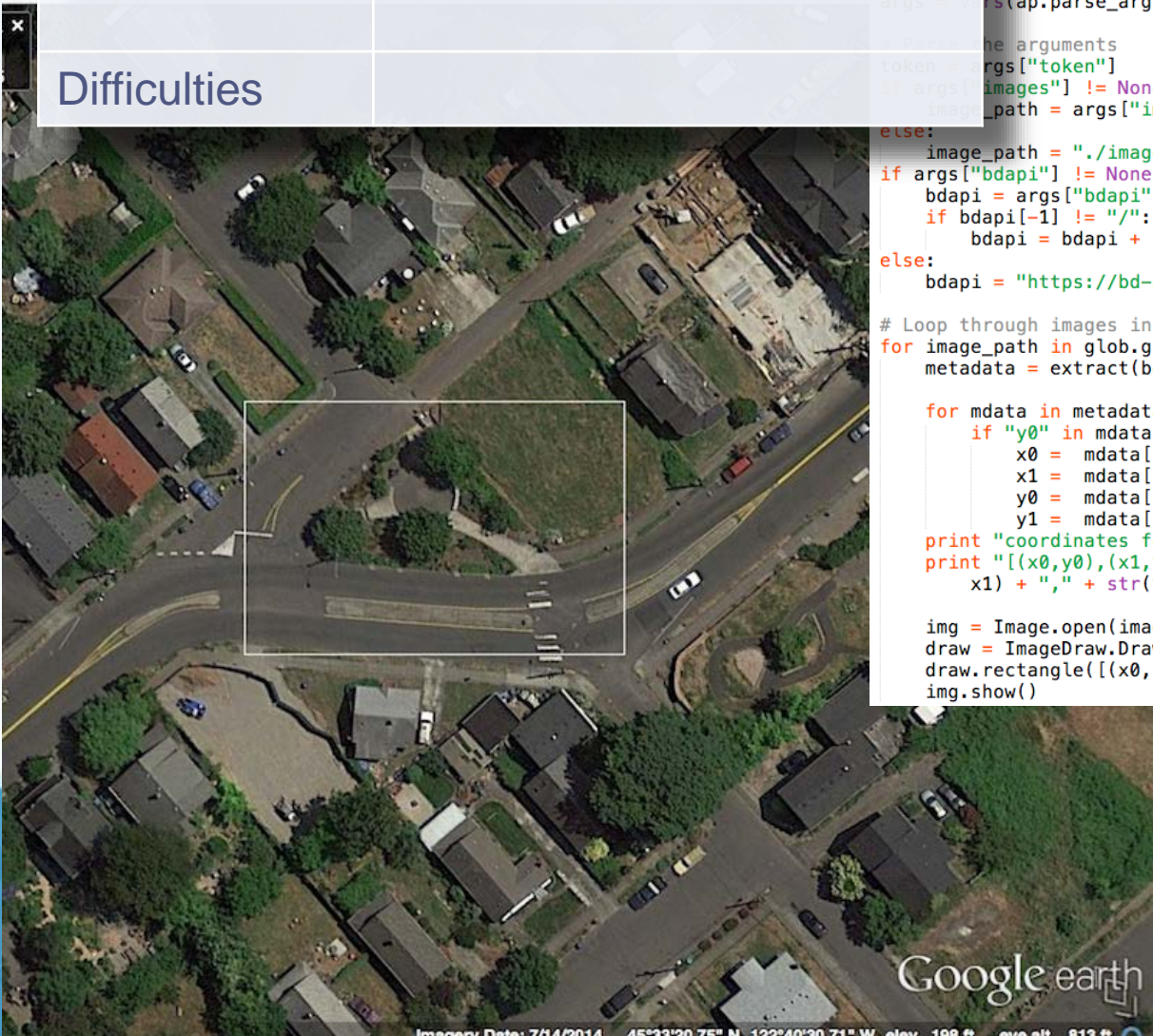
    for mdata in metadata:
        if "y0" in mdata["content"]:
            x0 = mdata["content"]["x0"]
            x1 = mdata["content"]["x1"]
            y0 = mdata["content"]["y0"]
            y1 = mdata["content"]["y1"]
            print "coordinates for bioswale bounding box in" + image_path
            print "[(x0,y0),(x1,y1)] = [{" + str(x0) + "," + str(y0) + "}, {" + str(x1) + "," + str(y1) + "}]

    img = Image.open(image_path)
    draw = ImageDraw.Draw(img)
    draw.rectangle([(x0, y0), (x1, y1)])
    img.show()

```



Total Code from 1 File



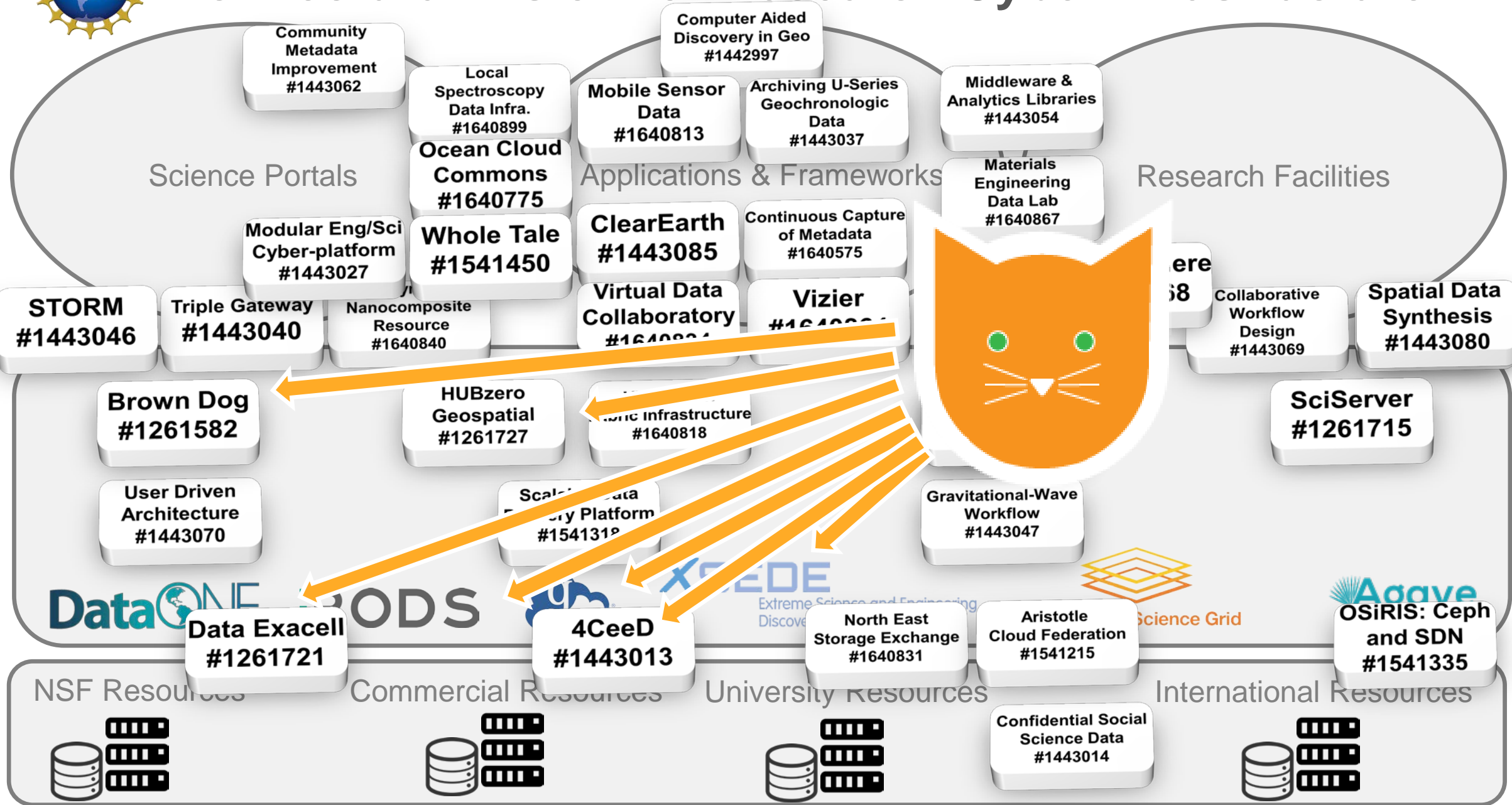


Architectural Vision for Research Cyberinfrastructure

Discipline Specific Environments

Integrative Services

Resources



Clowder (2013-Present)

NSF Innovative Systems and Software: Applications to NARA Research Problems (OCI-0525308)



The screenshot shows the Clowder homepage in a browser window. The address bar displays <https://clowder.ncsa.illinois.edu/clowder/>. The page features a blue header with the Clowder logo, navigation links for 'Explore' and 'Help', a search bar, and buttons for 'Sign up' and 'Login'. The main content area includes a 'Welcome to Clowder' section with a paragraph of text: 'Welcome to Clowder, a scalable data repository where you can share, organize and analyze data. This is a demo instance to try the system out. Please do not use this instance to store real data. We delete the content of this instance when we need to and it does not have very much disk space available. Thank you.' To the right of the main content is a 'Resources' sidebar with links for 'Spaces', 'Collections', 'Datasets', 'Files', 'Bytes', and 'Users'. At the bottom left, it says 'Powered by Clowder (1.3.2#16 branch:master sha1:b7a81e8)'.

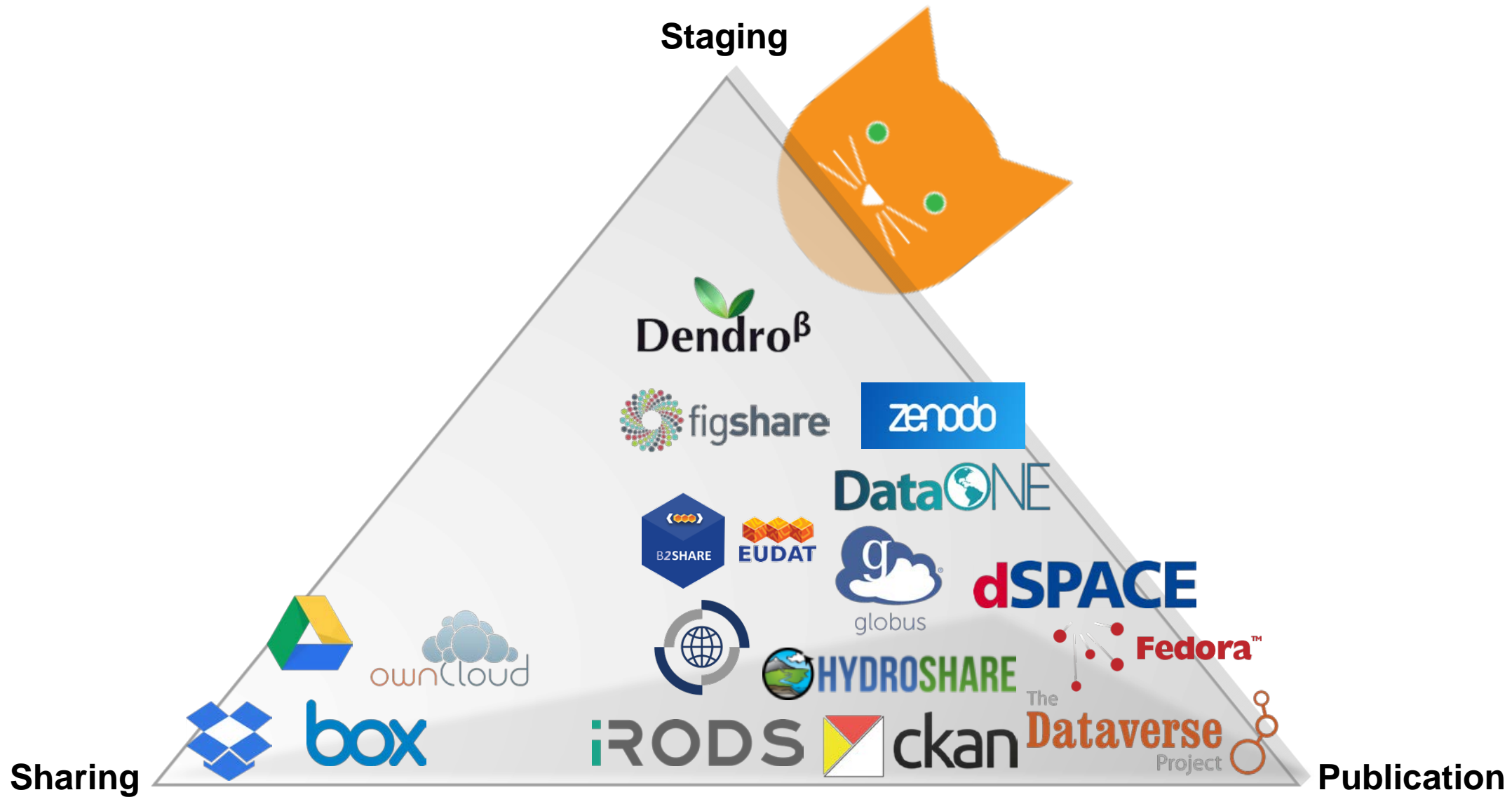
The screenshot shows a file page in Clowder. The browser address bar shows <https://clowder.ncsa.illinois.edu/clowder/files/5a1f1481e4b0cfb1ad158e1f?dataset=5a1d7faae4b0...>. The file name is '6011937799.jpg'. The main content area displays a thumbnail of an aerial photograph of a residential street. To the right of the image is a metadata panel with the following information: 'Type: image/jpeg', 'File size: 494.8 kB', 'Uploaded on: Nov 29, 2017 14:11:45', 'Uploaded by: Bardia Heidari Haratmeh', and 'Status: PROCESSED'. Below this is a 'License' section with 'Type: All Rights Reserved' and 'Holder: Bardia Heidari Haratmeh'. Further down is a 'Dataset containing the file' section with a link to 'GI identification dataset'. At the bottom, there is a 'Tags' section with a tag 'Bioswale Detected [Section]' and a 'Tag' button.

- **Active curation (AC)** involves recording data and metadata as close to the source as practical and driving that acquisition through the deployment of capabilities that help data producers manage their research.
- **Social Curation (SC)** drives this economic analysis further, looking at ways that crossgroup interactions can further motivate best practices.

Machine Learning

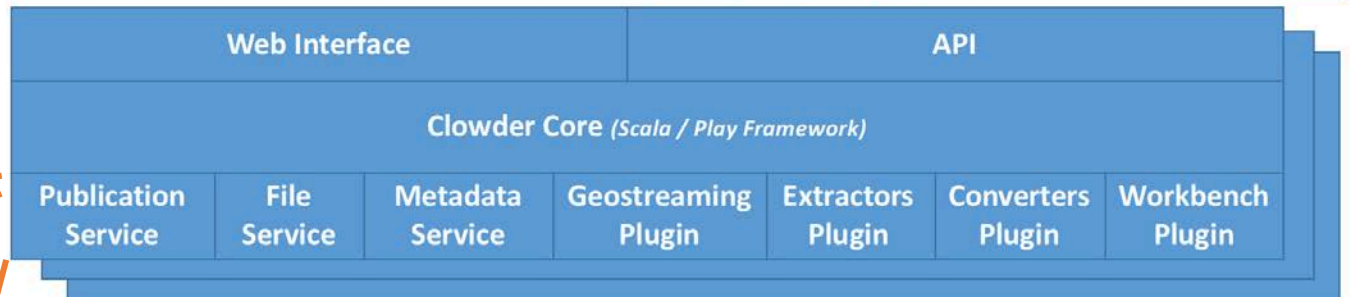
J. Myers and M. Hedstrom, "Active and Social Curation: Keys to Data Service Sustainability," NDS Consortium Planning Workshop, 2014

<http://sead-data.net/sites/default/files/pubs/ActiveandSocialCurationKeystoDataServiceSustainability.pdf>

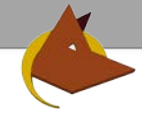
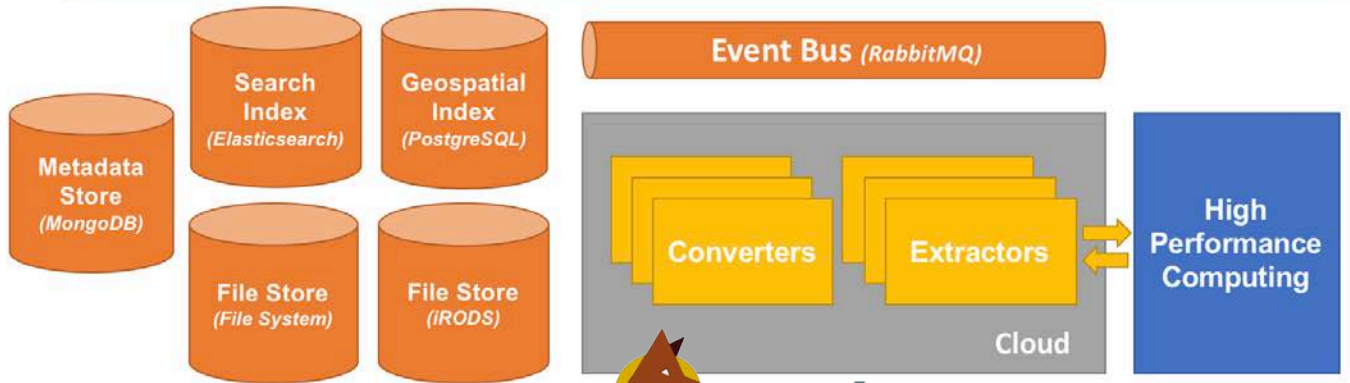




Clients
Services



IDEALS
ICPSR
DataONE
ckan
MATERIALS DATA FACILITY
HYDROSHARE



Broad Data & Tool Support

BD Tool Catalog BD Fiddle Contribute Admin Search

CN8452_POD6_11-20-13.tiff

in computer-aid diagnosis

Created by Anonymous User
Created on Apr 08, 2015

Created by DW WF 2015-04-08T11:39:58

Geospatial Layers

Metadata

- Extracted by http://clowder.ncsa.illinois.edu/extra
- content: Tubular Cast: 103

Deployments: [bd-api](#) [bda-api-dev](#)

Submitter: mchenry

Last modified : Nov 20, 2016

Clowder Spaces

452 Z. Ge et al.

2.4 Carbon Storage of P Periods

Carbon storage of the ecosystem of carbon storage of vegetation layer, tree layer has a distinct carbon storage of the ecosystem thinning periods from large to 5 years after thinning. The thinning periods from large to 5 years after thinning. Carbon storage of the ecosystem thinning periods from large to 5 years after thinning. Carbon storage of the ecosystem thinning periods from large to 5 years after thinning. Carbon storage of the ecosystem thinning periods from large to 5 years after thinning.

Time after thinning (a)	Carbon stocks (t/ha)		
	Vegetation layer	Soil layer	Ecosystem
CK	28.54 (66.65)	14.28 (33.35)	42.82 (100)
5	26.97 (67.60)	12.93 (32.40)	39.90 (100)
10	40.43 (75.68)	12.99 (24.32)	53.42 (100)

5. Numbers in the brackets were the percentage of carbon stocks in different layers.

Cluster Modes

- Any tile
- CHA of the q the memory k
- Memory in sa quadrant as C

Sub-NUMA-Clus Divide chip into domains, c.f., mu system. SNC-4 to a 4-socket no

November 22, 2017

Download Delete Follow

Clowder

Collections Datasets Files Tags API

Person Tracking

Filename: simple_video_1.mp4
Type: video/mp4
Uploaded on: Jun 23, 2015 17:18:45
Uploaded by: Sandeep Satheesan

License
Type: All Rights Reserved
Holder: Sandeep Satheesan

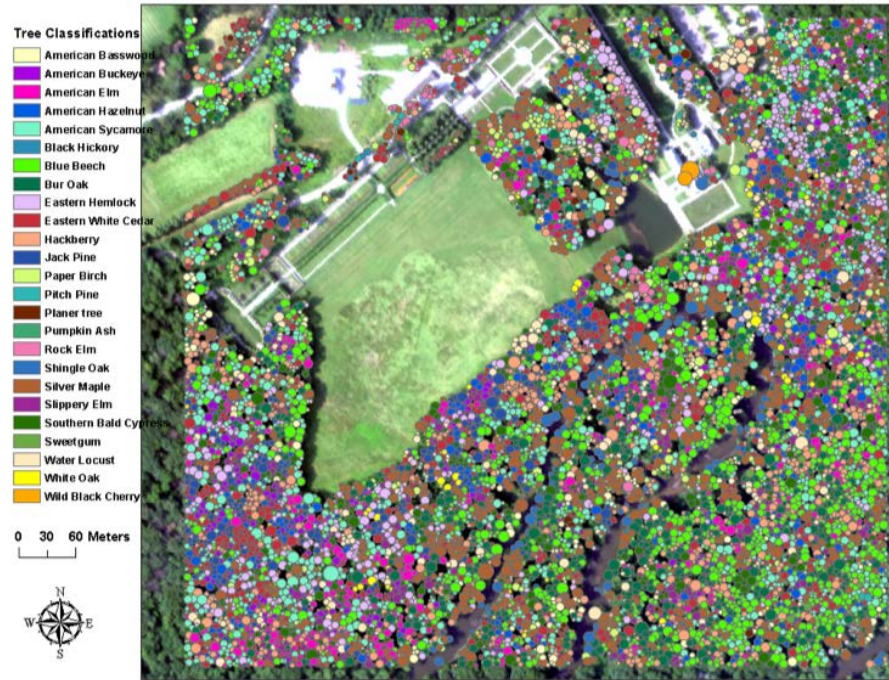
41.07 fps
anim: 0
texture: 1
sort: 0
render: 1
#Fris: 477306
#PPris: 1429069
#Crash: 21

Distance: 41591

Enter annotation description
Annotation A

OK Cancel annotation addition

Broad Data & Tool Support



Terra Ref Collections Datasets Files Tags

Thumbnail Thumbnail Thumbnail Thumbnail Thumbnail

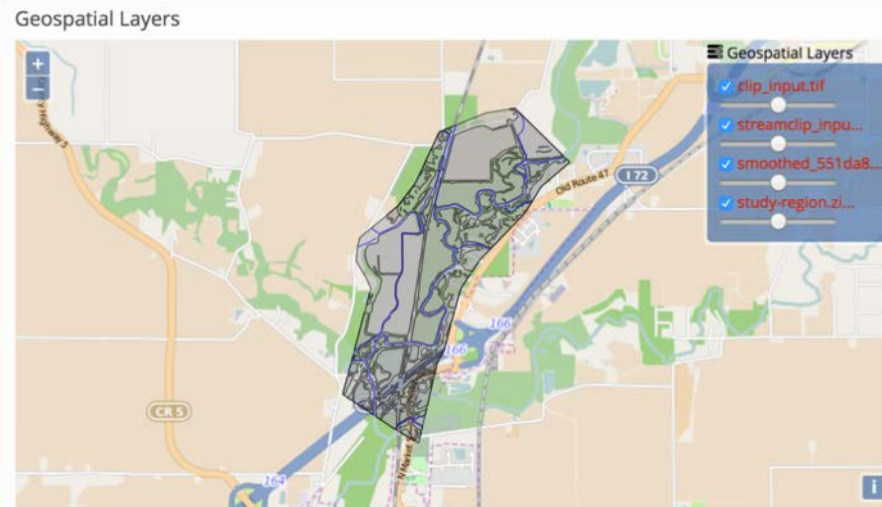
Filename: VIS_SV_270_2700_379478.jpg
Type: image/jpeg
Uploaded on: Nov 10, 2015 15:12:15
Uploaded by: Yan Liu

License
Type: All Rights Reserved
Holder: Yan Liu
Edit

Tags
• plantcv [File]
• image analysis [File]

Datasets containing the file
yantest Detach Add

Sensors Associated with this file



Metadata

Metadata Definitions

The following metadata definitions will be available throughout Clowder.

Label <i>i</i>	URI <i>i</i>	Type
Abstract	http://purl.org/dc/terms/abstract	String
Alternative Title	http://purl.org/dc/terms/alternative	String
Audience	http://purl.org/dc/terms/audience	String
CSDMS Standard Name	http://csdms.colorado.edu/wiki/CSN_Searchable_List	List
ODM2 Variable Name	http://vocabulary.odm2.org/variablename	List
References	http://purl.org/dc/terms/references	String
SAS Spatial Geocode	http://ecgs.ncsa.illinois.edu/gsis/sas/geocode	Location
SAS Variable Name	http://ecgs.ncsa.illinois.edu/gsis/sas/vars	Queryable List

Metadata

Add metadata

ODM2 Variable Name

ODM2 Variable Name

Select field

- wa
- waterLevel**
- radiationIncomingLongwave
- waterUsePublicSupply
- TDRWaveformRelativeLength
- waterPotential
- radiationNetLongwave
- waterVaporDensity
- waterDepth
- waterVaporConcentration
- waterUseCommercialIndustrial

Submit

Added by Luigi

Location Name

Latitude: 4

Longitude:

02 Choose a dataset... [what's this?](#)

Existing Datasets

New Dataset

Basic Load Custom

Load a template Clear

Choose a dataset template:

polyvillic nanoparticles

Choose a name for your dataset:

Example... Sample Name, PECVD Oxide, Diffusion

Add

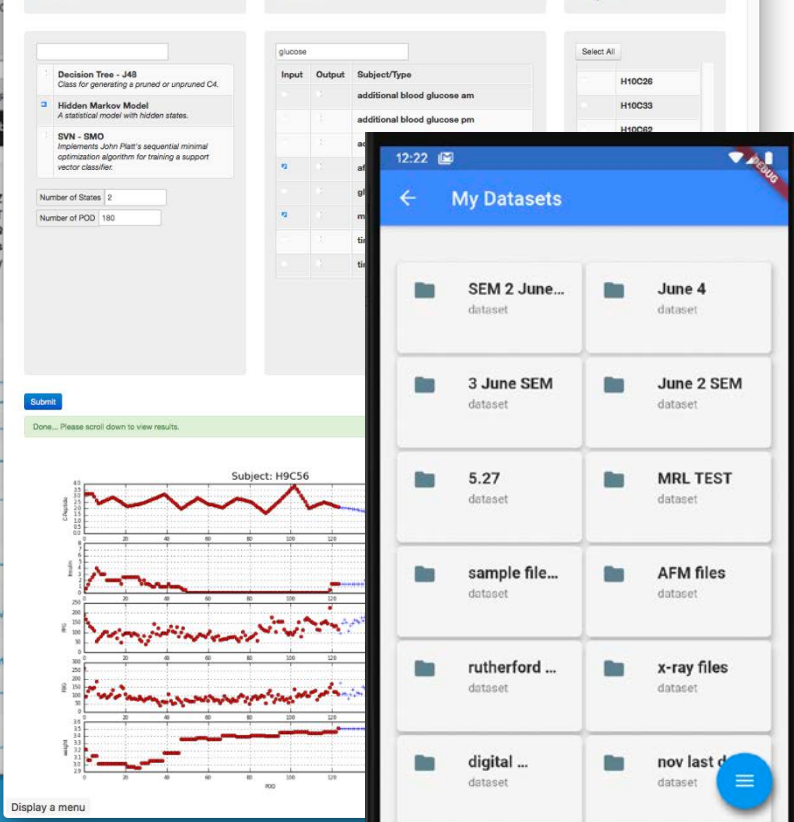
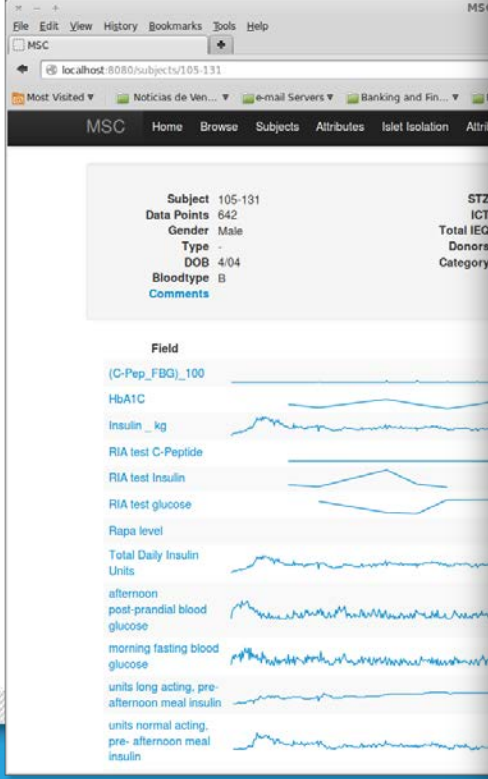
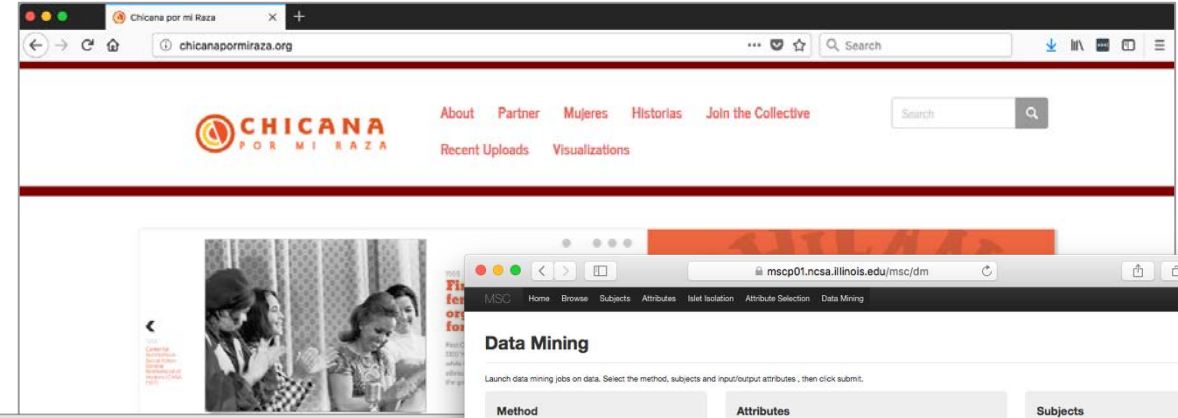
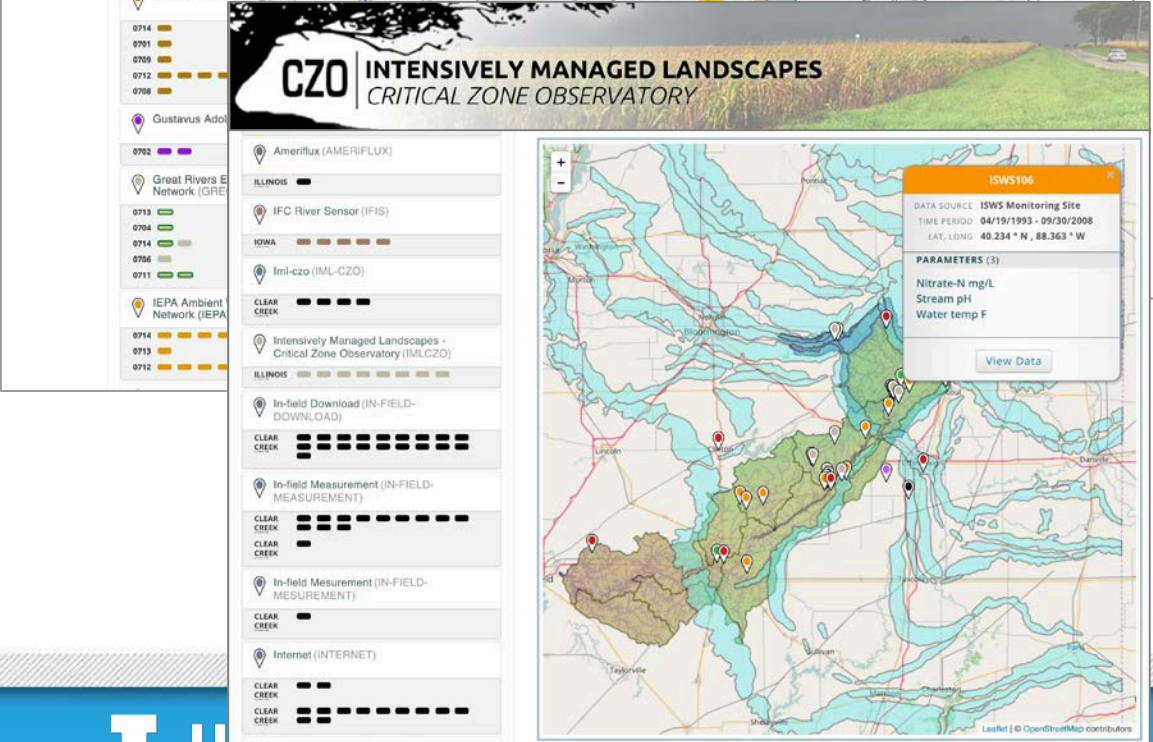
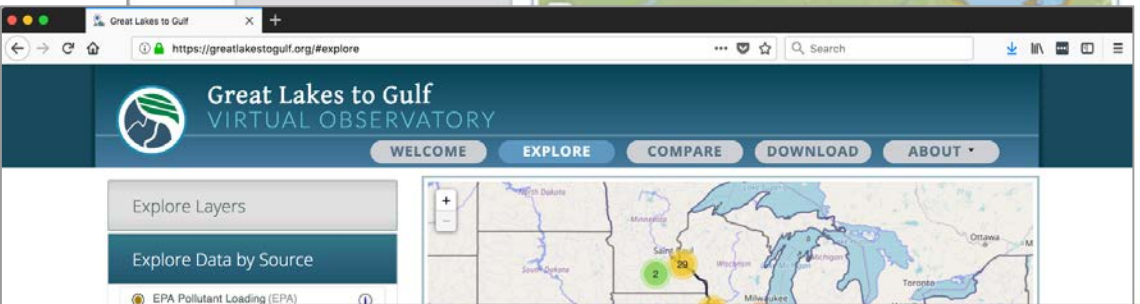
Name: Value: Remove

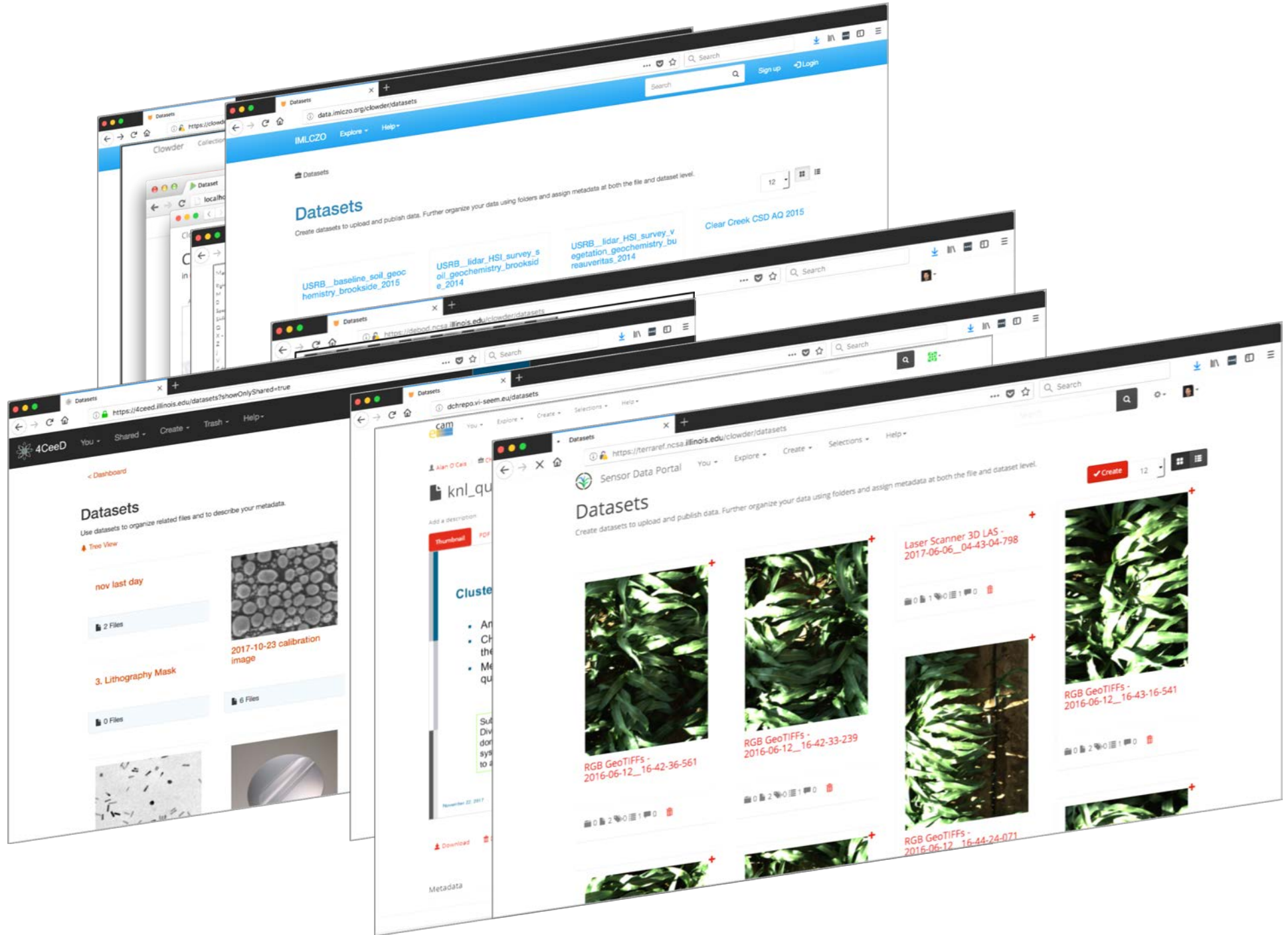
Name: Value: Remove

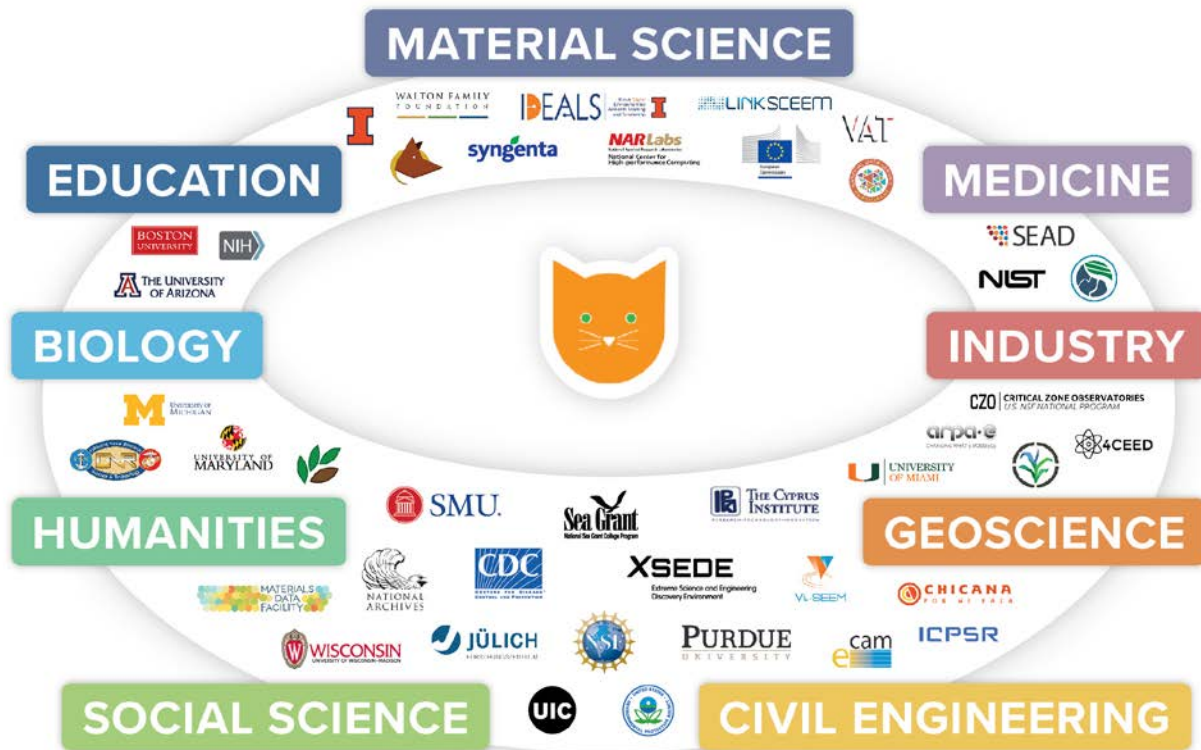
Incubation1 time (min) : Remove

Create Dataset

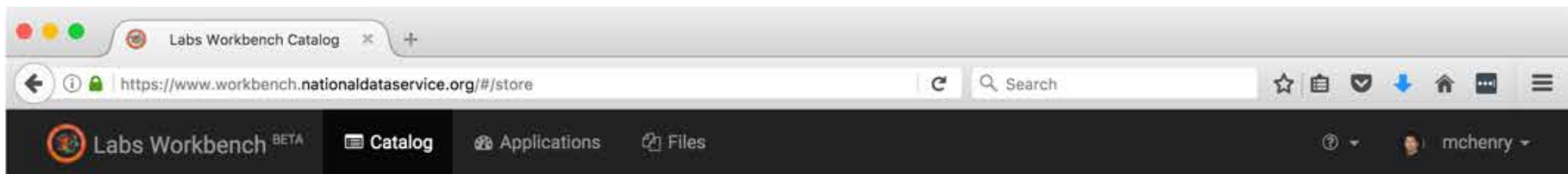
Custom Community Dashboards







		Biology	Civil Engineering	Comp. and Inf. Science	Education	Geoscience	Humanities	Industry	Materials Science	Medicine	Social Science
Brown Dog	NSF	X	X	X		X					X
BRACELET	NSF			X					X		
Chicanapormiraza	Univeristy of Michigan						X				
Countermeasures against Radiation	BARDA	X								X	
Crowd-Sourced Green Infrastructure	NSF		X								X
DataNet SEAD	NSF, NDS	X	X			X					
EarthCube GeoSemantics	NSF					X					
eCam	European Commission				X						
IML-CZO	NSF	X			X	X					
Immunomodulatory Effects of MSC	NIH									X	
Great Lakes Monitoring	Illinois-Indiana SeaGrant					X					
Great Lakes to Gulf	NGRECC, Walton Foundation					X					
Groupscope	NSF										X
KISTI	KISTI	X		X							
LinkSCEEM	European Commission					X	X				
NARA	NSF			X							
NIST	NIST								X		
PecAn	NSF	X			X	X					
RIVEEL3D	Cyprus Institute						X				
SIMDAS	European Commission						X				
Southern Methodist University	SMU		X	X							
Syngenta	Syngenta	X						X			
T2-C2	NSF			X					X		
Taiwan NCHC	NCHC			X						X	
TERRA-REF	ARPA-E	X			X	X					
TRECC	ONR				X						
Vector-Borne Disease	CDC	X									
Vi-SEEM	European Commission						X				
XSEDE Decomposing Bodies	NSF			X			X				
XSEDE Image Analysis of Rural Photography	NSF			X			X				
XSEDE Large Scale Video Analytics	NSF			X			X				
XSEDE Real Stories of Bad Kids	NSF			X			X				X



Catalog

Labs Workbench / Catalog

Filter applications... Import Create

Choose an application below to configure

Analysis Toolbox ?
Customizable server for analysis tools
toolmanager
Add

CKAN ?
A tool for making open data websites to help manage and publish collections of data.
Archive Repository Python Data management ckan
Add

Cloud9 C/C++ ?
C/C++ development environment based on Cloud9
Development environment C/C++ cloud9cpp
Add

Cloud9 Docker ?
Docker environment based on Cloud9
Development environment Docker cloud9docker
Add

Cloud9 GO ?
Go development environment based on Cloud9
Development environment Go cloud9go
Add













Cloud9 Java ?
Java development environment based on Cloud9
Development environment Java cloud9java
Add

Cloud9 Node.js ?
Node.js development environment based on Cloud9
Development environment Javascript cloud9nodejs
Add

Cloud9 PHP ?
PHP development environment based on Cloud9
Development environment PHP cloud9php
Add

Cloud9 Python ?
Python development environment based on Cloud9
Development environment Python cloud9py
Add



 <h3>Clowder</h3> <p>A scalable data repository where you can share, organize and analyze data.</p> <p>Active data management Scala Data management clowder</p> <p>View 1 Add</p>	 <h3>Dataverse</h3> <p>A web application to share, preserve, cite, explore and analyze research data</p> <p>Archive Repository Java Data management dataverse</p> <p>View 1 Add</p>	 <h3>Docker</h3> <p>Docker command-line environment for NDS Labs</p> <p>Development environment Go Docker docker</p> <p>Add</p>
 <h3>DSpace</h3> <p>Turnkey software used for creating open access digital repositories.</p> <p>Repository Java dspace</p> <p>Add</p>	 <h3>ElasticSearch</h3> <p>A distributed search and analytics engine</p> <p>Search engine elasticsearch2</p> <p>Add</p>	 <h3>File Manager</h3> <p>Cloud Commander file manager and editor</p> <p>Javascript Data transfer cloudcmd</p> <p>Add</p>
 <h3>Girder</h3> <p>Web-based data management platform.</p> <p>Python Active data management Data management girder</p> <p>Add</p>	 <h3>HTTP Tunnel</h3> <p>NDS Labs Utility for accessing TCP-based services via secure tunnel</p> <p>System utilities Go chisel</p> <p>Add</p>	 <h3>iQvoc</h3> <p>SKOS-based vocabulary management system</p> <p>Ruby Semantic iqvoc</p> <p>Add</p>
 <h3>iRODS Cloudbrowser</h3> <p>Web-based interface to any iRODS grid</p>	 <h3>iRODS iCAT</h3> <p>open source data management</p>	 <h3>Jenkins CI</h3> <p>An open-source automation server</p>

Status	Name	ID	Console	Edit	Remove	Help
Stopped	Dataverse	s6n4g2-dataverse	>_			
Stopped	PostgreSQL	s6n4g2-postgres	>_			
Stopped	Solr	s6n4g2-solr	>_			
Stopped	Rserve	s6n4g2-rserve	>_			

HTTP Basic Authentication Disabled

DSpace

Status	Name	ID	Console	Edit	Remove	Help
Stopped	DSpace	szca0w-dspace	>_			
Stopped	PostgreSQL	szca0w-postgres	>_			

HTTP Basic Authentication Disabled

Root Dataverse > Account

Edit Account

- My Data
- Notifications
- Account Information

Here are all the dataverses, datasets, and files you have access to. You can filter through them by publication status and roles.

Search my data...

- Dataverses**
- Datasets**
- Files**

Publication Status

- Published**
- Unpublished**
- Draft**
- In Review**
- Deaccessioned**

Roles

- Admin**
- File Downloader**
- Dataverse + Dataset Creator**
- Dataverse Creator**
- Dataset Creator**
- Contributor**
- Curator**
- Member**

Sorry, nothing was found for these roles: Admin, File Downloader, Dataverse + Dataset Creator, Dataverse Creator, Dataset Creator, Contributor, Curator, Member



Configuration for Dataverse

ID: s6n4g2-dataverse

Docker

Registry	Image Name	Image Tag
docker.io	ndslabs/dataverse	Default (latest)

Environment

Property	Value	Default
ADMIN_PASSWORD	vuW994bZdR	
MAIL_SERVER	smtp.ncsa.illinois.edu	smtp.ncsa.illinois.edu
POSTGRES_DB		
POSTGRES_PASSWORD		
POSTGRES_USER		
RSERVE_PASSWORD	rserve	rserve
RSERVE_USER	rserve	rserve
TOOLMANAGER_URL		
TWORAVENS_URL	https://\$(NDSLABS_STACK)-tworavens.\$(NDSLABS_DOMAIN)	https://\$(NDSLABS_STACK)tworavens.\$(NDSLABS_DOM

Endpoints

Protocol	Internal Address	External Address
HTTP	10.254.79.87:8080	s6n4g2-dataverse.workbench.nationaldatasevice.org

Dataverse

Status	Name
✓ Ready	Dataverse
✓ Ready	PostgreSQL
✓ Ready	Solr
✓ Ready	Rserve

DSpace

Status	Name
⏻ Stopped	DSpace
⏻ Stopped	PostgreSQL

Application Disabled Remove Launch

Logs Help

Shutdown

Remove Help

Kenton McHenry

Labs Workbench / Applications

Clowder

Status	Name	ID	Console	Edit	Remove	Help
Stopped	Clowder	so150m-clowder	>_			
Stopped	MongoDB	so150m-mongo	>_			
Stopped	RabbitMQ	so150m-rabbitmq	>_			
Stopped	PlantCV extractor	so150m-plantcv	>_			

HTTP Basic Authentication Disabled

Dataverse

DSpace

Data Curation Workbench *BETA* | Catalog | Applications | Files | willis@

Catalog

Data Curation Workbench / Catalog

Filter applications...

Choose an application below to configure

- CKAN**: A tool for making open data websites to help manage and publish collections of data.
 - Archive | Repository | Python
 - Data management
 - View | Add
- CKAN DataStore**: Postgres database.
 - RDMS | Add
- Dataverse**: A web application to share, preserve, cite, explore and analyze research data.
 - Archive | Repository | Java
 - Data management
 - Add
- File Manager**: Cloud Commander file manager and editor.
 - Javascript | Data transfer | cloudcmd
 - View | Add
- PostgreSQL Studio**: Web based interface PostgreSQL.
 - Add

www.ischool.ndslabs.org | University of Illinois. Powered by Labs Workbench.

Analysis Workbench *BETA* | Catalog | Applications | Files | mchenry@

Catalog

Analysis Workbench / Catalog

Filter applications...

Choose an application below to configure

- Hyperspectral/NCO**: Hyperspectral workflow development environment.
 - Add
- Jupyter/NetCDF4**: Jupyter SciPy Notebook with NetCDF support.
 - Development environment | Javascript | Python
 - Analysis environment
 - Add
- Jupyter/PlantCV**: Jupyter SciPy Notebook with PlantCV support.
 - Development environment | Javascript | Python
 - Analysis environment
 - View | Add
- JupyterLab for pyCloudier**: Extensible computational.
 - Add
- RStudio Geospatial**: IDE for the R programming language.
 - Development environment | Analysis environment
 - Add

Powered by Labs Workbench.

Workbench *BETA* | Catalog | Applications | Files | mchenry@

Catalog

Workbench / Catalog

Filter applications...

Choose an application below to configure

- File Manager**: Cloud Commander file manager and editor.
 - cloudcmd
 - Add
- Jupyter/Einstein Toolkit**: Jupyter SciPy Notebook with Einstein Toolkit dependencies.
 - Development environment | Javascript | Python
 - Analysis environment
 - Add

UI v1.0.13-d1vel
API v1.0.12 2017-07-27 20:04

Labs Workbench *BETA* | Applications | Catalog | Files | cherny@

Catalog

Labs Workbench / Catalog

Filter applications...

Choose an application below to configure

- Dataverse**: A web application to share, preserve, cite, explore and analyze research data.
 - Archive | Repository | Java | Data management
 - View | Add
- IRODS Cloudbrowser**: Web-based interface to any IRODS grid.
 - Data transfer | Java | cloudbrowser
 - View | Add
- Active data**:
 - View
- PostgreSQL**: Relational database platform.
 - RDMS | postgres
 - Add
- TwoRavens**: A collection of statistical tools for data exploration, analysis, and meta-analysis.
 - Javascript | Analysis environment | tworavens
 - Add

All rights reserved. © 2017 Board of Trustees of the University of Illinois. Powered by Labs Workbench.

UI v1.0.7 2016-10-24 12:59
API v1.0.7 2016-10-24 18:18

Labs Workbench *BETA* | Catalog | Applications | Files | cherny@

Catalog

Labs Workbench / Applications / Catalog

Filter applications...

Choose an application below to configure

- Cloud9 C/C++**: C/C++ development environment based on Cloud9.
 - Development environment | C/C++ | cloud9
 - Add
- Cloud9 Docker**: Docker environment based on Cloud9.
 - Development environment | Docker | cloud9
 - Add
- Cloud9 GO**: Go development environment based on Cloud9.
 - Development environment | Go | cloud9
 - Add
- Cloud9 Java**: Java development environment based on Cloud9.
 - Development environment | Java | cloud9
 - Add
- Cloud9 Node.js**: Node.js development environment based on Cloud9.
 - Development environment | Javascript | cloud9
 - Add
- Cloud9 PHP**:
 - Add
- Cloud9 Python**: Python development environment based on Cloud9.
 - Development environment | Python | cloud9
 - Add
- ElasticSearch**: A distributed search and analytics engine.
 - Search engine | elasticsearch
 - Add
- Cloud9 IDE**: General-purpose editor with python, git, ssh support.
 - Development environment | Python | cloud9
 - View | Add
- File Manager**: Cloud Commander for manager and editor.
 - Javascript | Data transfer | cloudcmd
 - Add
- Globus Connect**: Launch a Globus Personal Endpoint.
 - Development environment | Javascript | Python
 - Analysis environment
 - Add
- Jupyter**: Jupyter SciPy Notebook with Python and various kernels, netCDF, etc. dependencies.
 - Add
- PostgreSQL Studio**: Web-based interface to PostgreSQL, for access to BITable.
 - Add
- RStudio Server**: RStudio server with graphical interface and python. User management is multi-tenant.
 - Development environment | Analysis environment | R
 - Add
- Visualization tools**: Remote desktop environment with visualization tools including OSG and Paraview.
 - Add

All rights reserved. © 2017 Board of Trustees of the University of Illinois.

UI v1.1.0
API v1.1.0 2019-02-17 20:27

Labs Workbench *BETA* | Catalog | Applications | Files | mchenry@

Catalog

Labs Workbench / Catalog

Filter applications...

Choose an application below to configure

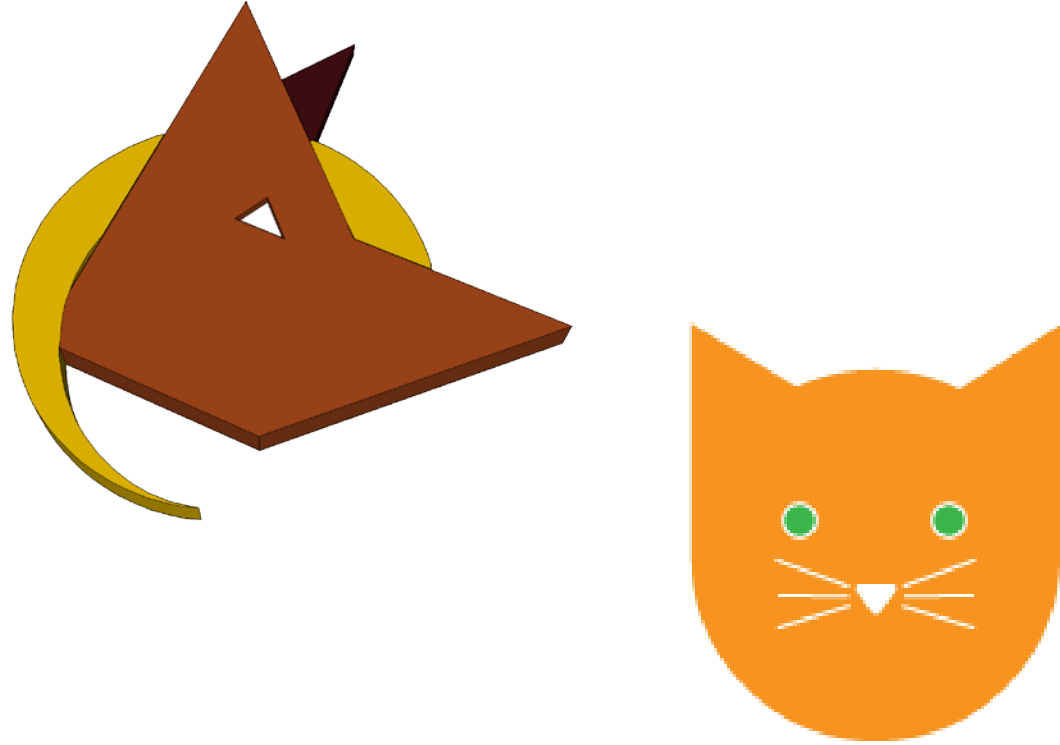
- CHORDS**: Easy-to-use system to acquire, navigate and distribute real-time data streams.
 - Building block | chords
 - Add
- File Manager**: Cloud Commander file manager and editor.
 - Javascript | Data transfer | cloudcmd
 - Add
- GeoDeepDive Development Environment**: Python development environment based on Cloud9.
 - Development environment | globus
 - Building block | Python
 - Add
- InfluxDB**: Scalable database for metrics, events, and real-time analytics.
 - influx
 - Add
- Jupyter Minimal Notebook**: Jupyter Minimal Notebook.
 - Javascript | Python | Analysis environment | jupyter
 - Add
- MongoDB**: A cross platform document-oriented NoSQL database.
 - NoSQL | mongo
 - Add



Further/Custom Analysis

The screenshot shows the 'Catalog' page in Analysis Workbench. The top navigation bar includes 'Analysis Workbench BETA', 'Catalog', 'Applications', and 'Files'. The user 'maxzilla' is logged in. The main content area is titled 'Catalog' and contains a search bar with the text 'Filter applications...'. Below the search bar, there are buttons for 'Import' and 'Create'. A central instruction reads 'Choose an application below to configure'. There are six application cards displayed in a grid:

- Cloud9 Python environment for TERRA**: Python development environment with pyClowder and TERRA dependencies. Tags: Development environment, Python. ID: cloud9terra. Buttons: View, Add.
- File Manager**: Cloud Commander file manager and editor. Tags: Javascript, Data transfer. ID: cloudcmd. Button: Add.
- Jupyter/TERRA**: Jupyter SciPy Notebook with TERRA dependencies. Tags: Development environment, Javascript, Python. ID: jupyterterra. Buttons: View, Add.
- PostgreSQL Studio**: Web-based interface to PostgreSQL. ID: pgstudio. Button: Add.
- RStudio Server**: RStudio Server with TERRA dependencies. Tags: Development environment, Analysis environment, R. ID: rstudio. Button: Add.
- Visualization tools**: Xpra container containing visualization tools. ID: xpraterra. Button: Add.



129,471,866 datasets/files transformed

Cost of the "Cloud"

W. Gropp, R. Harrison, M. Abbot, R. Grossman, P. Kogge, P. Raghavan, D. Reed, V. Taylor, K. Yelick, J. Eisenberg, S. Bradley, "Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020", National Academies of Sciences, Engineering, Medicine, 2017



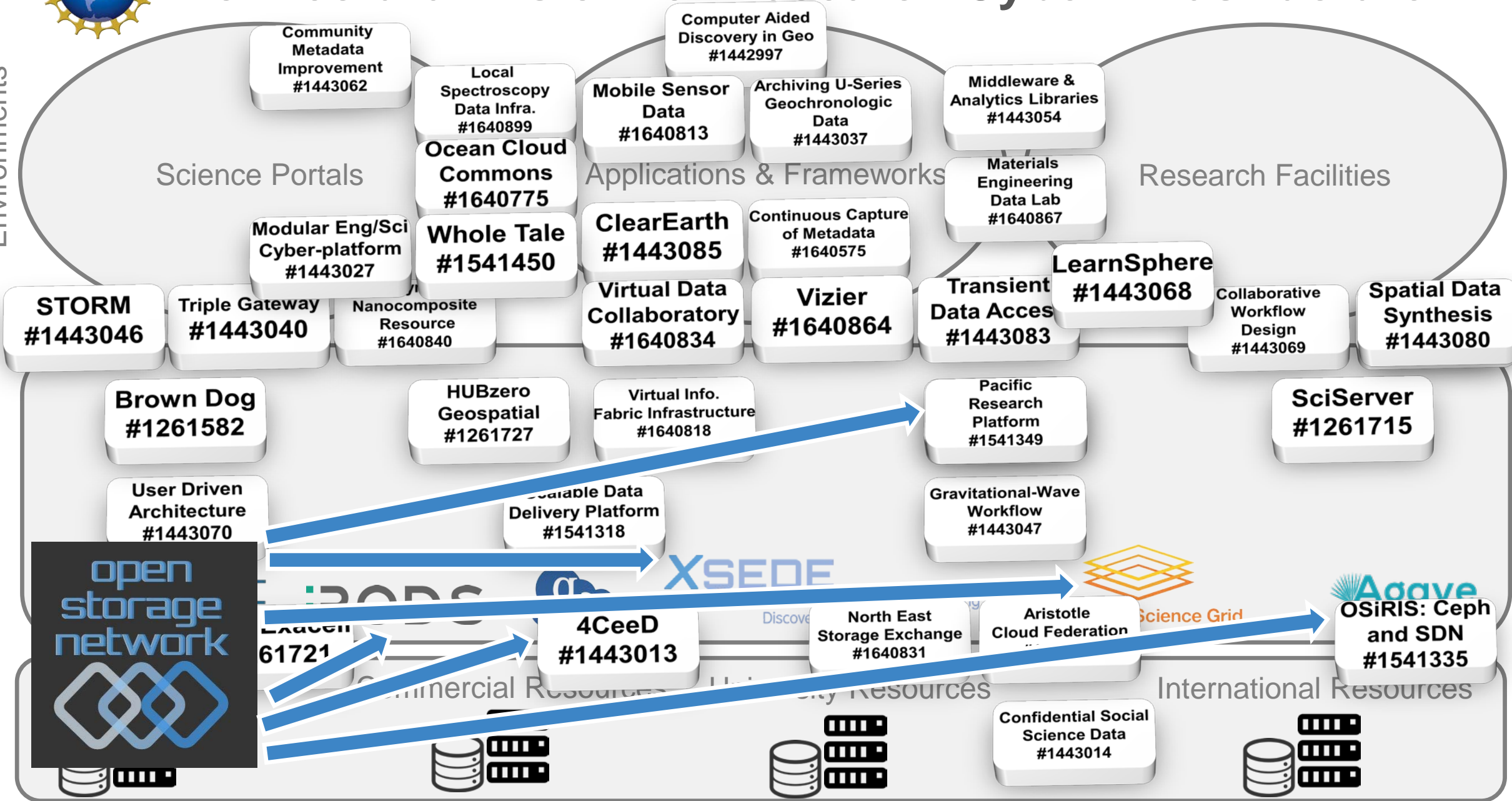


Architectural Vision for Research Cyberinfrastructure

Discipline Specific Environments

Integrative Services

Resources





CC*

Coordinated campus-level
cyberinfrastructure components of
data, networking, and computing
infrastructure across 200+ universities

Big Data Hubs
Stimulate regional grassroots
partnerships focused on Big Da



Hubs based on Census Regions of the
United States



Alaska & Hawaii are part of the West Region
US Territories can participate in any region

OSN Nodes

	88-Disk OSN Node	24 Disk OSN Node (128GB)	24 Disk OSN Node (256GB)	36 Disk OSN Node
Total Sockets	2	1	1	1
Per-CPU Cores (HT)	6 (12)	12 (24)	12 (24)	18 (36)
Total CPU Cores (HT)	12 (24)	12 (24)	12 (24)	18 (36)
CPU Speed	3.4 GHz	2.1 GHz	2.1 GHz	2.3 GHz
Total GFLOPS	652.8	403.2	403.2	662.4
Total RAM (GB)	256	128	256	256
RAM Speed (MHz)	2400	2400	2400	2666
Drive Count (AKA OSD)	88	24	24	36
Per-Node Data Storage (TB)	704	192	192	288
~1.5PB Node Count	2	8	8	5
Total Data Storage	1408	1536	1536	1440
Node Cost	\$51,492.00	\$15,528.00	\$16,672.00	\$17,845.00
Total Cost	\$102,984.00	\$124,224.00	\$133,376.00	\$89,225.00
\$ per TB	\$73.14	\$80.88	\$86.83	\$61.96
Utilized Rack Space (RU)	24	36	36	24
Max Power Consumption (Watts)	7600	9600	9600	6800
Ceph RAM:Storage Cap Ratio (Nearest to 1 is best)	0.36	0.67	1.33	0.89
Ceph CPU Thread: OSD Ratio (Nearest to 1 is best)	0.27	1	1	1



Data-Scope
Style Node



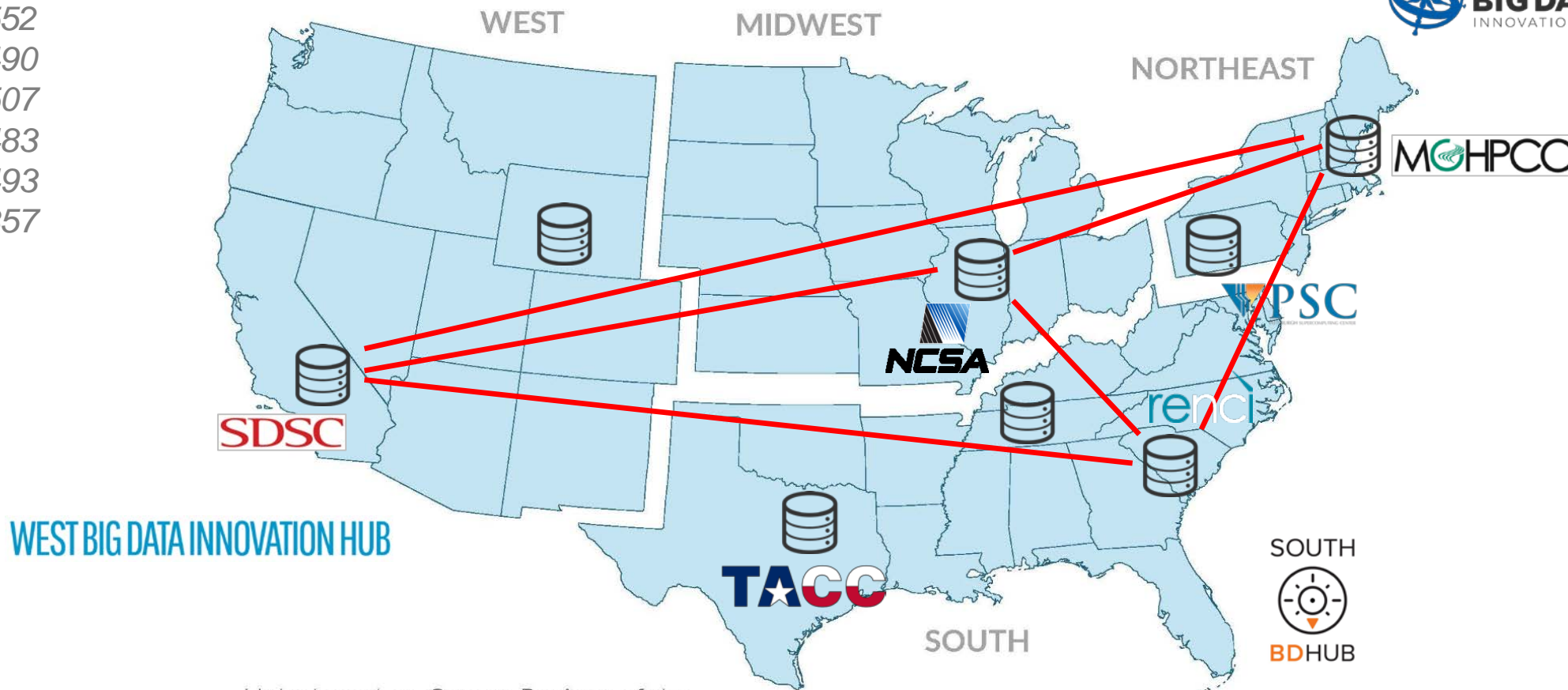


Open Storage Network

IIS-1747552
IIS-1747490
IIS-1747507
IIS-1747483
IIS-1747493
IIS-1836357

Midwest Big Data Hub

Accelerating the Big Data Innovation Ecosystem



WEST BIG DATA INNOVATION HUB

Hubs based on Census Regions of the United States

Alaska & Hawaii are part of the West Region
US Territories can participate in any region

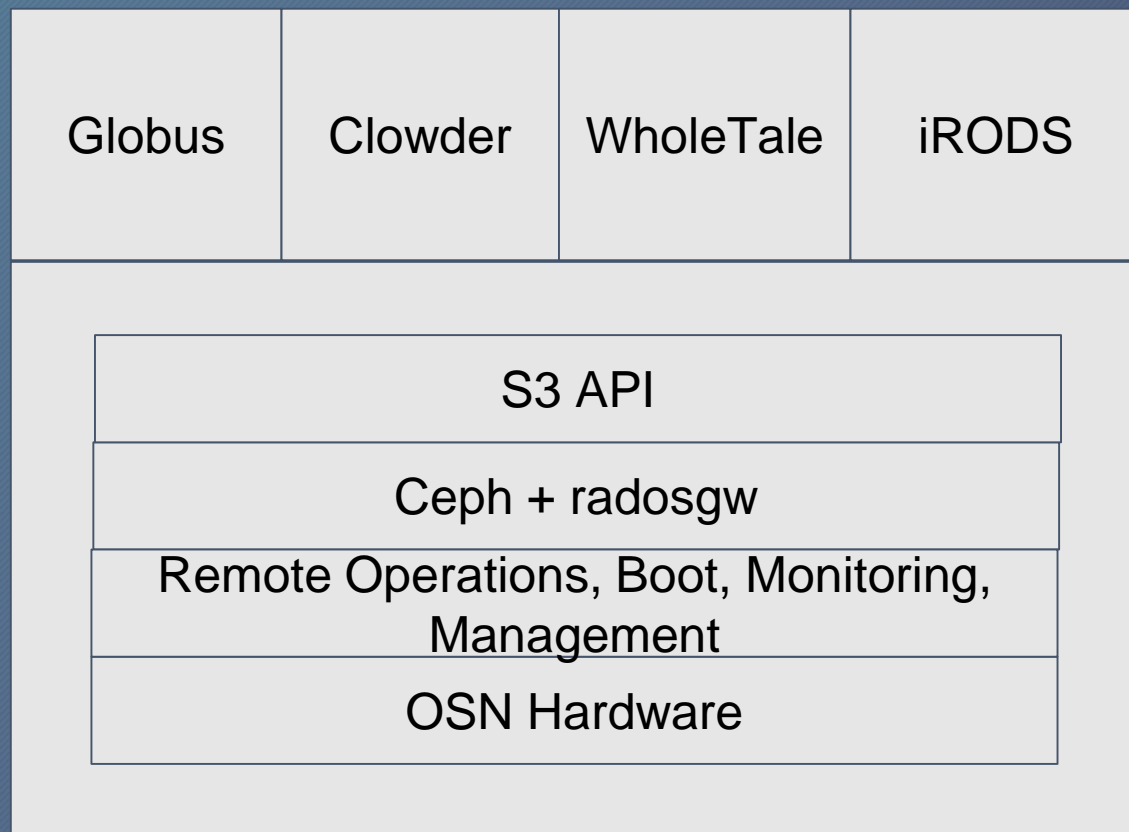
Use Cases

Project	Average size of data entities	Total data volume	Storage problem being solved	Use case
Critical Zone Observatories	10 MB	50 TB	Provide storage space and access to CZO datasets and community-generated data	Community long-tail data
TerraFusion	10 GB	1 PB	Transport datasets across the US at high speed, obtain data slices with high probability of reutilization	Experiment-to-site, Slice-and-compute
HathiTrust Research Center collection	200 MB	500 TB	Provide storage space and access to the HTRC dataset and further community-generated derivatives	Common resource access
Machine Learning	10 GB	1 PB	Make available a well-curated dataset for testing machine learning algorithms	Dataset-as-benchmark
Large Synoptic Survey Telescope	2 TB	100 PB	Transport datasets across the US at high speed, obtain data slices with high probability of reutilization, facilitate inter-site data processing	Experiment-to-site, Slice-and-compute, Workflow staging space
Combined Array for Research in Millimeter Astronomy	50 MB	50 TB	Transport datasets across the US at high speed, obtain data slices with high probability of reutilization	Experiment-to-site, Slice-and-compute

Architecture

Leverage NSF
Data Software
Investments
*e.g DataNet,
DIBBs, CSSI*

OSN Pod



Leverage NSF
XSEDE
...Capabilities



OSN Clowder Interface

- IML-CZO use case data



kubernetes



A screenshot of a web browser displaying the OSN Clowder interface. The browser tab is titled "OSN IML CZO" and the address bar shows the URL "osn.clowderframework.org/spaces/5d9e11cee4b0a90763505...". The page header includes "OSN Clowder" and navigation menus for "You", "Explore", "Create", "Selections", and "Help". The main content area is titled "OSN IML CZO" and includes a description: "Intensively Managed Landscapes Critical Zone Observatory. Intensively managed landscapes encompass agricultural, urban, and natural environments where human modifications of the landscapes intersect with the natural legacies of the land. The services on this page provide access to data managed by the IML-CZO." Below the description are action buttons: "Delete", "Create Dataset", "Create Collections", and "Search". The "Datasets in the Space" section shows "Viewing most recent datasets" and a "View All Datasets" link. Two dataset cards are visible: "Flux Tower Yearly Aggregates" with a description "Data aggregates based on data sent over Loggernet" and "Flux Tower Site Visit" with a description "Site visit on May 17th, 2016." On the right side, there are management options: "Manage Users", "Edit Space", "Manage Metadata Terms & Definitions", "Extractors", and "Follow". An "External Links" section at the bottom right contains the URL "http://data.imlczoo.org/".