



SCIENCE-AS-A-SERVICE WITH THE AGAVE PLATFORM

Rion Dooley @deardooley
dooley@tacc.utexas.edu

TACC AT A GLANCE



Personnel

160 Full time staff (~70 PhD)

Facilities

10 MW Data center capacity

New office facility to be completed
by start of 2016

Systems and Services

A Billion compute hours per year

5 Billion files, 50 Petabytes of Data,
Hundreds of Public Datasets

Capacity & Services

HPC, HTC, Visualization, Large scale
data storage, Cloud computing
Consulting, Curation and analysis,
Code optimization, Portals and
Gateways, Web service APIs, Training
and Outreach



EXTREME SCALE SUPERCOMPUTING



Stampede

- #7 HPC system in the world for computation 500k CPU core 9.7 PF

Lonestar 5

- Texas-focused Cray XC40 30,000 Intel Haswell cores 1.25 PF

Wrangler

- 0.6 PB usable DSSD flash storage w 1 TB/s read rate + 10 PB Lustre

Maverick

- 132 Fat nodes w dual 10 core Ivy Bridge + NVIDIA Kepler K40 GPGPU

Chameleon & Jetstream Cloud

- 1400 nodes OpenStack

Disk and Tape Storage

- 100+ PB storage in HIPAA-aligned data center



Coming Soon: Hikari

- Green computing system partnership with NEDO and NTT. 10k Haswell cores. HVDC and Solar (partial)
- Support for container ecosystem

TACC SUPPORTS AN INCREDIBLE AMOUNT & DIVERSITY OF RESEARCH

- Since 2013...
 - Over *2 Billion* processor hours delivered to end users
 - 6+ **million** successful jobs
 - About 10,000 students, faculty, and staff use our Stampede directly
 - Over 30,000 more use it indirectly via portals and services
 - Peer-reviewed requests for time (via XSEDE) run ~500% available hours
- **Stampede alone** supports nearly 2,500 funded projects across the United States and abroad

THE EVOLUTION OF A CYBERINFRASTRUCTURE



Once upon a time, most of us built garage-style clusters...



HPC systems have grown up since then and become much more powerful and sophisticated

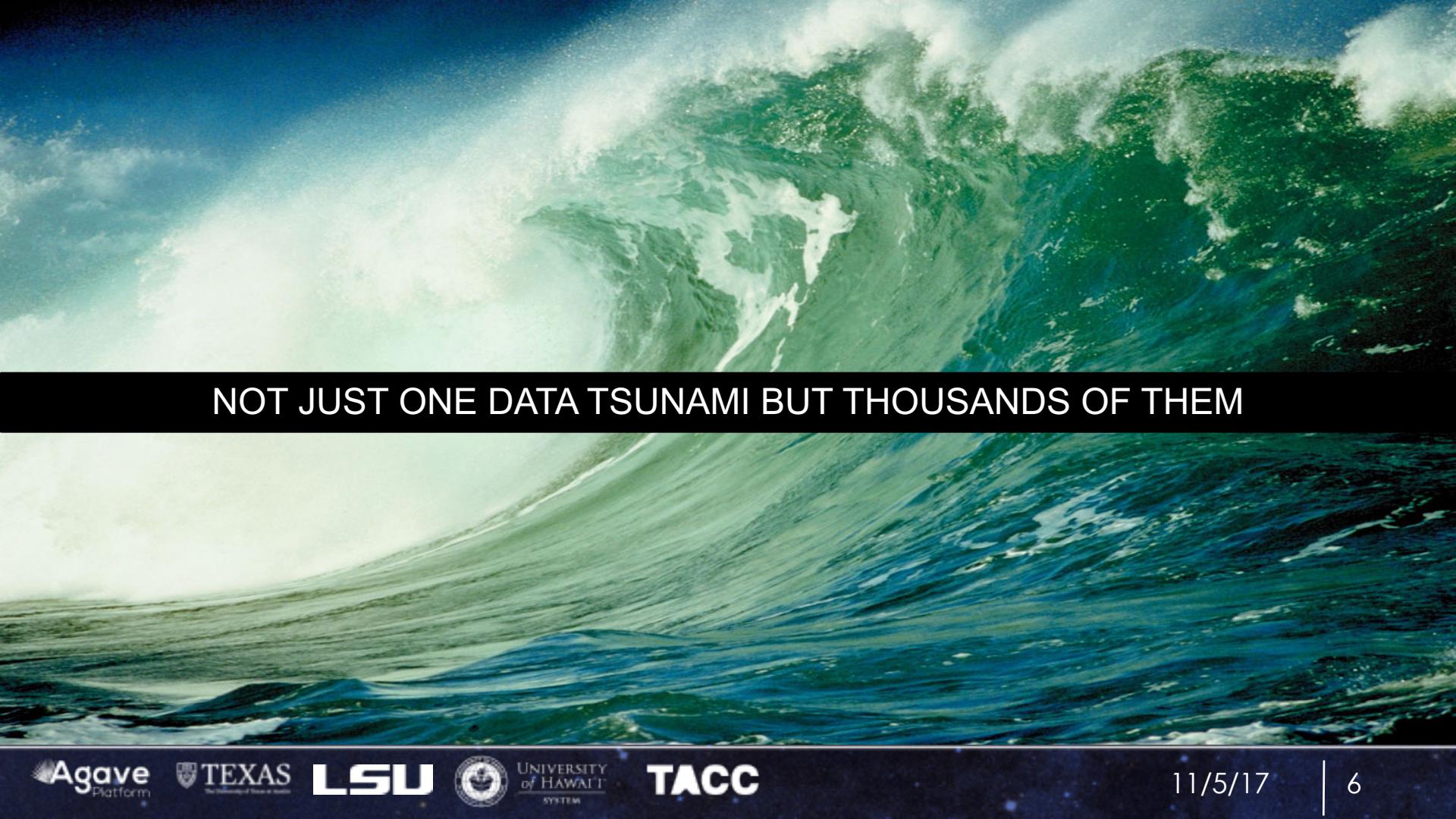


But a curious thing happened while HPC was running its victory lap...

We built a very successful center based on 3 pillars in our first 7 years

- *Simulation & HPC*
- *Visualization*
- *People (Consulting & Algorithms)*





NOT JUST ONE DATA TSUNAMI BUT THOUSANDS OF THEM

CAMBRIAN EXPLOSION OF LANGUAGE, TECHNOLOGY, AND HARDWARE

CATALYSTS

- Triumph of Linux and FOSS
- Rise of collaborative massive scale
- Emergence of Internet-scale technologies
- Rapid innovation to solve IO bound problems
- Repeal of Moore's Law

WORKFLOWS NOW TECHNICALLY COMPLICATED

LANGUAGES

- Python 2 & 3
- R
- Julia
- Perl
- Matlab
- Java
- Scala, Clojure, etc
- .NET
- C++
- Swift
- Haskell
- Go
- Javascript

FRAMEWORKS

- MapReduce: Hadoop, Storm, Pachyderm
- Event & Streaming: Kinesis, Azure Stream Analytics, Camel, Streambase
- Deep/Machine Learning: Watson, Azure BI, Tensorflow
- In-memory parsing: Kognito, Apache Spark
- New data warehouse: Snowflake
- Containers: Docker, Rocket, MESOS, Kubernetes
- Cloud: AWS, GCE, OpenStack, VMWare

HARDWARE

- Rise of many-core computing means 50-100 threads/node*
- Xeon / Xeon Phi
- GPU
- OpenPower
- ARM
- Multi-level memory architectures
- Hierarchical storage architectures
- FPGAs

DIVERSE DISTRIBUTED RESEARCH TEAMS



Mike

- Computing novice
- Works remotely at partner site



Eliza

- Masters specific analysis skills
- Readily adopts new tech



Paulo

- Staff computational expert
- Supports multiple projects



Nikolaidas Group

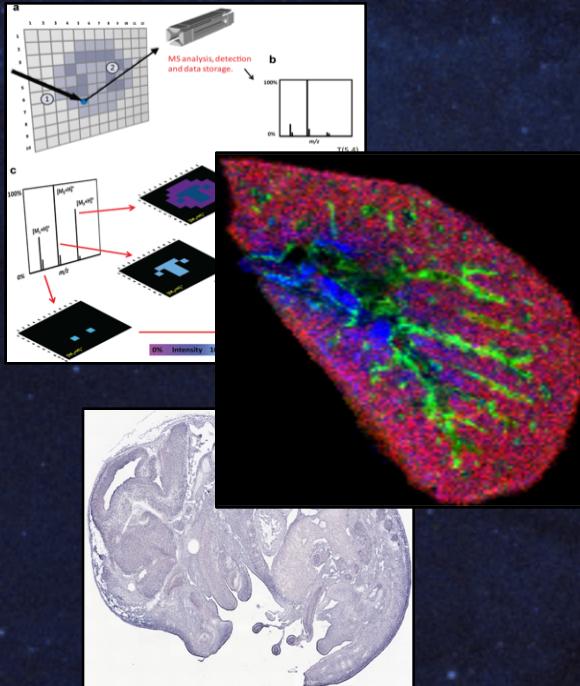
- Mostly experimentalists
- Strict data sharing & access



Roshan

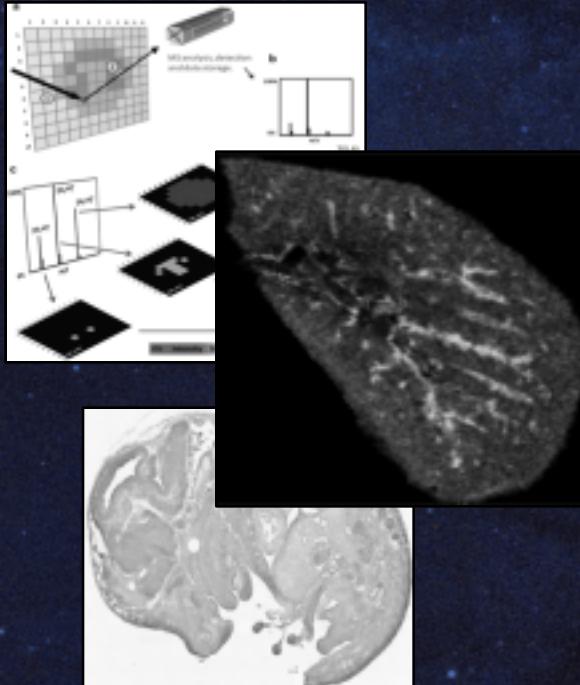
- Computationally experienced
- Focused on interpretation

THEIR NEEDS (30,000 FT. VIEW)



- ▶ Store, organize, share *primary data*
- ▶ Iteratively perform 1' analyses
- ▶ Store, organize, share *derived data products*
- ▶ Iteratively generate and explore hypotheses
- ▶ Share analytical code with the scientific public
- ▶ Integrate results from new experiments
- ▶ Publish data alongside plots, visualizations and analytical tools

THEIR NEEDS (500 FT. VIEW)

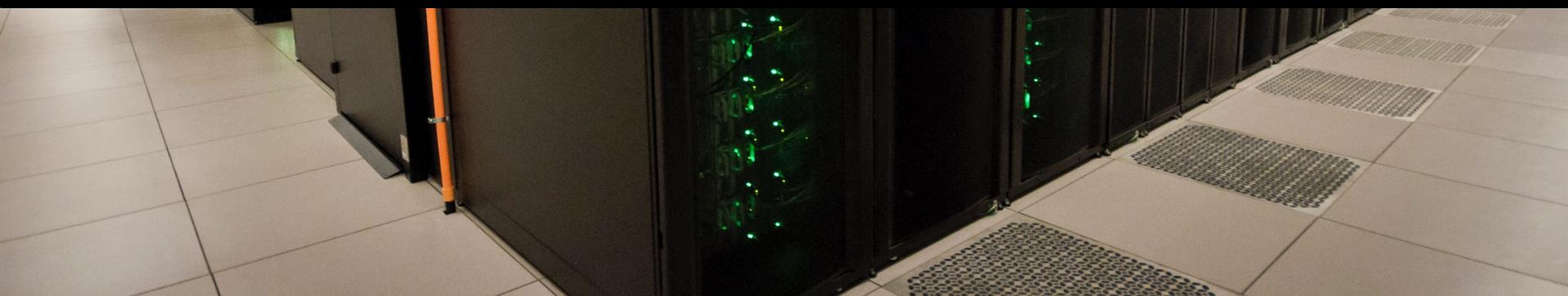


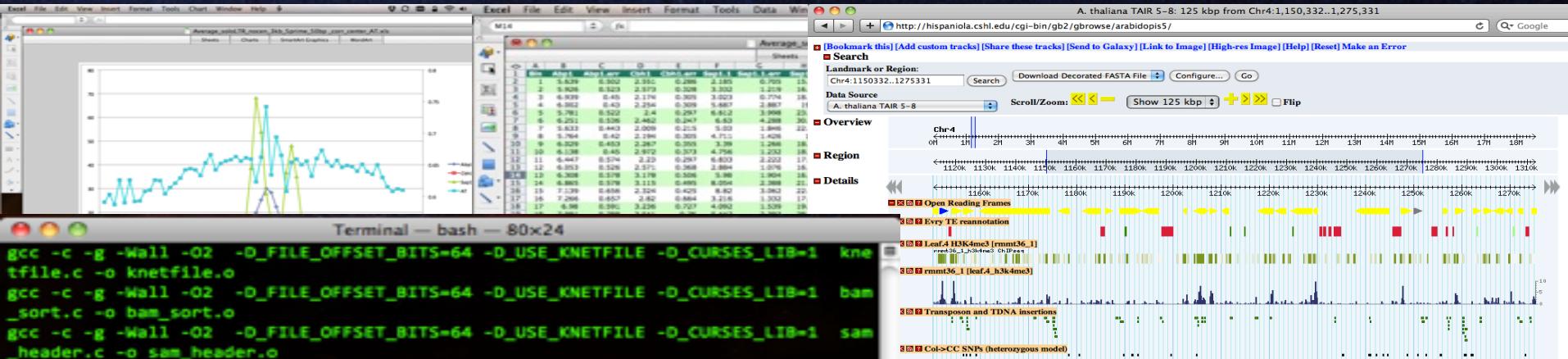
- ▶ Data lifecycle management
 - ▶ Fine-grained permission management
 - ▶ Discoverability
 - ▶ Version control
- ▶ Domesticating promising new analysis codes based on often immature technology
- ▶ Doing reproducible computational science
- ▶ Adopting efficient analytical methods

HOW DO WE HELP RESEARCHERS WITH SUCH DIVERSE NEEDS AND BACKGROUNDS?



BUILD A MASSIVE STORAGE CLOUD NEXT TO INNOVATIVE, POWERFUL, USABLE COMPUTERS AT THE END OF FAST INTERNET PIPES





MANY DOMAIN SCIENTISTS ARE NOT EXPERTS AT COMPUTING TECHNOLOGY.
CREATE PURPOSE-BUILT, HIGHLY INTUITIVE INTERFACES



DESIGNSAFE.CI (Top Left): A dashboard for managing files on Stampede, showing a file list and upload progress for 'Telem.tcl'.

API Explorer (Top Right): A screenshot of the Agave API Explorer interface, showing a list of available APIs for the Thalassine API.

Stampede (Bottom Left): A mobile-style interface for monitoring Stampede system status, showing utilization at 89% and a queue of 6 other users.

VDJ SERVER (Bottom Right): A screenshot of the VDJ SERVER interface, showing a list of projects and their details, including a 'Job Output Project' and 'Sample_01' data entries.

Texas Pandemic Flu Simulator: A screenshot of the Agave Platform interface, specifically the 'Community Data' section for the Thalassine API. It shows a grid of service icons and descriptions, such as 'agv! airport_generator_alignmentbag version 0.2' and 'ebi! interact_cytoscape_by_locus version 0.3'.

Point-and-click interfaces

- Data management, sharing, and analysis
- Publishing reproducible analysis workflows
- Discovery of new or updated tools and data
- Interactive visualization of results

Backed by world-class computing and data capacity

R version 3.0.3 (2014-03-06) -- "Warm Puppy"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: x86_64-unknown-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English environment

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help,
'start()' for an HTML browser interface to R, and
Type 'q()' to quit R.

> |

A Genomics Example Using the pyspark Library

In this example, we make use of the pyspark library to determine if any pathogens are present in a sample. The basic idea is to make use of k-mers, a biological analog of n-grams, to compute the "distance" from a known pathogen genome to the DNA in our sample. We can use different metrics for the distance, as will be shown below.

We note that this is a python3 notebook. At the moment, python3 is required to use the pyspark library.

```
In [1]: import string, os
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline
from IPython.display import Image, display, Math, Latex, SVG, HTML
import numpy as np
from scipy.cluster.hierarchy import linkage,dendrogram
from scipy.spatial.distance import pdist
# from urllib2 import urlopen
from urllib.request import urlopen
import pyspark
sc = pyspark.SparkContext('local[*]')
```

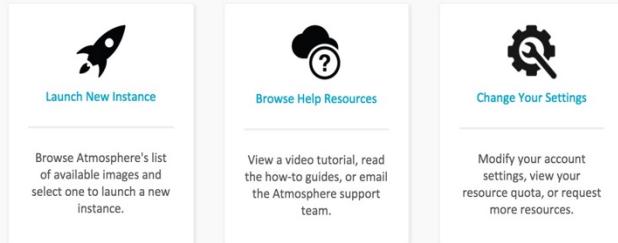
Pathogens

Hosted SaaS

- JupyterHub notebooks
- Rstudio
- Web-based VNC

Also, backed by world-class computing and data capacity

Getting Started



Resources in Use Need more?



Instance History (5 instances launched)



Updated a few seconds ago

Easy to use Cloud Computing

- Atmosphere (Cyverse)
- Jetstream (IU,UA,TACC)
- Chameleon (UC,TACC)

Cloud consoles are aimed at sysadmins and unintuitive.

We're changing that with improved UX and support

- APIs are still available
- No cost to end user



GIVE EXPERTS ACCESS TO EVERY SINGLE ONE OF YOUR BUILDING BLOCKS.
WEB SERVICE APIs EVERYWHERE. AUGMENT WITH PROFESSIONAL TOOLING.

MIKE: LEARNING COMPUTATIONAL SKILLS



Mike generates mass-spec data from his samples, which are used to decide on future experiments

- ▶ Eliza and Paulo have published analytical scripts for his data as 'apps' into a project web portal
- ▶ Paulo has wired up the lab's equipment to send data directly to remote storage
- ▶ Mike can perform basic analysis and reporting on his data from the web interface
- ▶ Roshan can see Mike's results in the project portal and discuss them with him
- ▶ Mike collaborates with Eliza to improve the results of their analytical scripts

ELIZA: AUGMENTING HER RESEARCH CAPABILITY



Eliza collects and analyzes *in situ* hybridization imagery as a major part of her inquiry

- ▶ She uses code developed by Paulo and others to perform feature detection
- ▶ She helps Paulo develop code for Mike to use in his proteomics work
- ▶ She has developed scripted workflows that run locally on her laptop to complete her analyses
- ▶ She writes her own code in Python and R for aggregate analysis and shares it via the Docker Hub and the project portal
- ▶ She presents her data viz via interactive Rshiny apps that she deploys to the project portal

PAULO: CONCENTRATING ON COMPUTATIONAL RESEARCH



Paulo likes to solve hard computational problems but is also responsible for lab research infrastructure

- ▶ He has automated data movement from instruments to remote storage, including duplication to AWS Glacier. Roshan gets the bill ☺
- ▶ He builds and maintains the project web portal. He didn't have much experience with such technology when the project started
- ▶ Paulo has used Spark to developed new software for feature extraction from *in situ* hybridization images – he can automatically deploy it to the portal, powered by TACC Wrangler, as part of his build process
- ▶ He is working on a paper describing his software and is getting valuable feedback from other folks he has shared it with via the project portal and public source repositories

ROSHAN: FOCUSING ON THE BIG PICTURE



- ▶ Roshan has deep experience in gene expression analysis - Paulo has populated the project portal with many of the tools she needs to accomplish her goals
- ▶ She collaborates with Mike and Eliza on interpreting their experimental results
- ▶ Because Eliza has included a custom notification in her scripting workflow, Roshan knows when new image analyses have been completed and can schedule time to look at them
- ▶ She can work with Paulo to enable the project web portal to make use of her newly-awarded XSEDE computing and storage allocation
- ▶ She routinely shares results with experimentalist colleagues and reviewers via a simple Dropbox-like interface



Agave

Platform

<https://github.com/agaveplatform/sydney-2017-training>

AGAVE IS A MULTI-TENANT PAAS
DELIVERING SCIENCE-AS-A-SERVICE SOLUTIONS
IN HYBRID CLOUD ENVIRONMENTS

-

Think of it like



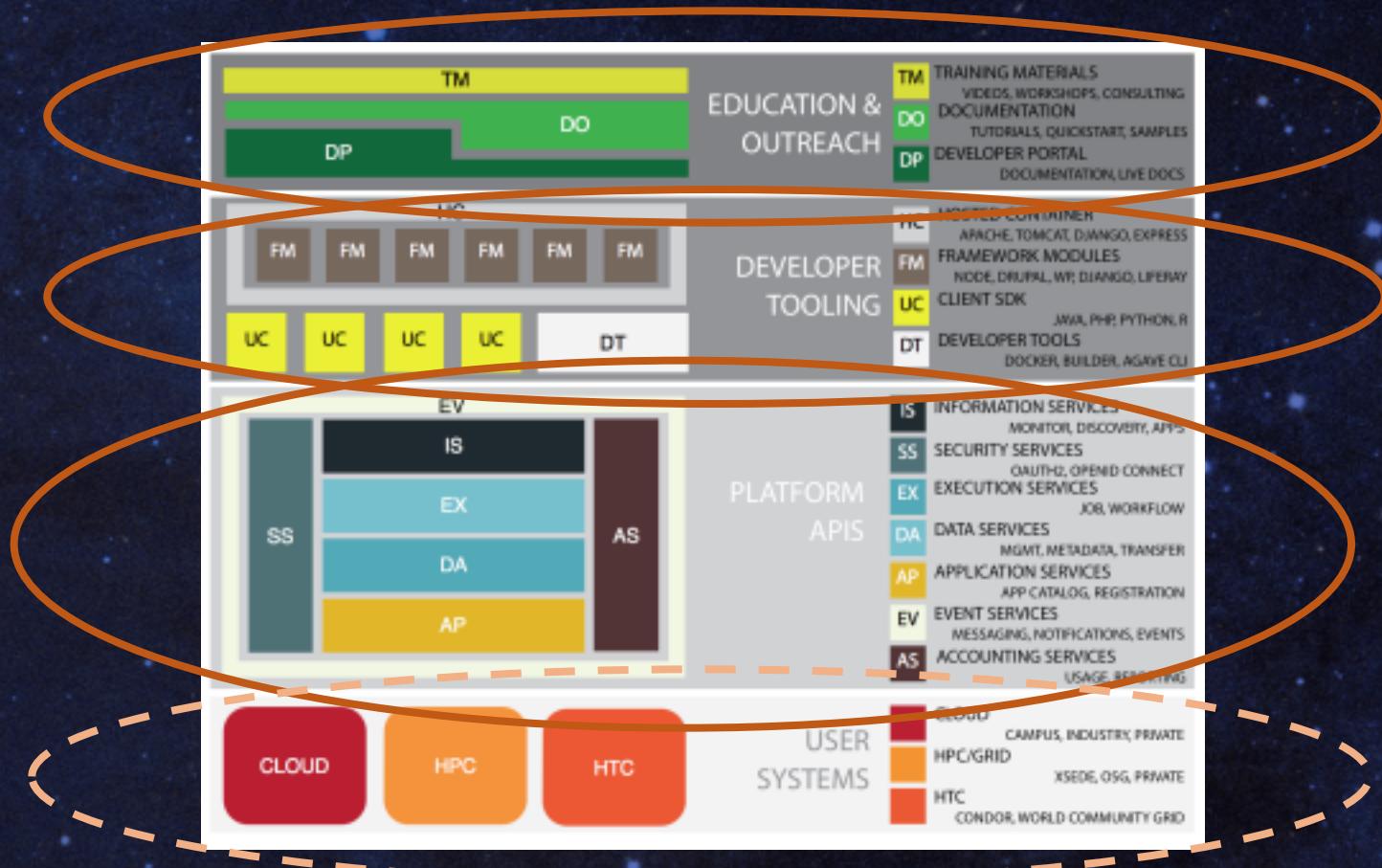
salesforce

for Science

WHAT IS AGAVE?

- ▶ Works with your new and legacy inf
- ▶ On premise, hybrid, or hosted deployment
- ▶ Multi-tenant
- ▶ Secure by default





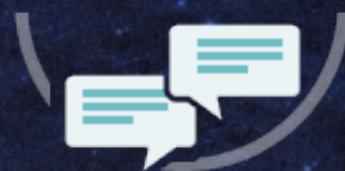
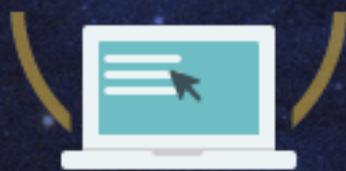
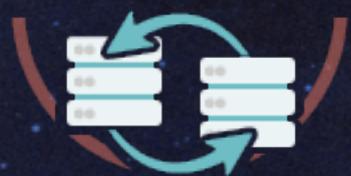
WHAT DOES IT DO?

MANAGE
DATA

RUN
CODE

COLLABORATE
ANYWHERE

CONNECT
ANYTHING



WHAT CAN AGAVE DO FOR ME?





Agave

Platform

PLATFORM : 30,000 FOOT VIEW

- ▶ Expand the reach of your infrastructure
- ▶ Enable new channels of collaboration
- ▶ Increase your research velocity



PLATFORM : KEYBOARD VIEW

1. Identity-as-a-Service
2. Flexible Authentication
3. API Management
4. Admin Services



PLATFORM : KEYBOARD VIEW

Identity-as-a-Service

- ▶ Integrated with, or supplement your existing IaaS solutions
- ▶ OpenID Connect compliant
- ▶ Extensible user profiles.
- ▶ REST profile management API (OpenID, Schema.org, SCIM, OIDC Federation, vCard,...)
- ▶ REST group management API (SCIM)
- ▶ Fine grained user and group RBAC



PLATFORM : KEYBOARD VIEW

Flexible Authentication

- ▶ OAuth2
- ▶ Identity federation
- ▶ SAML/Shibboleth
- ▶ Single Sign-On
- ▶ MFA
- ▶ Custom OAuth flows
- ▶ Service Accounts
- ▶ User/group mapping



PLATFORM : KEYBOARD VIEW

API Management

- ▶ SSL termination
- ▶ Gateway proxy
- ▶ Load balancing
- ▶ JWT generation
- ▶ Lifecycle management
- ▶ RBAC per API
- ▶ Swagger 2/3 & OAI support
- ▶ Metrics & reporting
- ▶ Self-service client registration
- ▶ Self-service API registration
- ▶ Developer documentation



PLATFORM : KEYBOARD VIEW

Admin Services

- ▶ Service accounts
- ▶ Admin roles and role mapping
- ▶ Client management
- ▶ Provisioning and publishing
- ▶ Credential generation & impersonation
- ▶ Token validation
- ▶ Header inspection
- ▶ On-demand log streams
- ▶ Self-service API registration and management





Agave

Science APIs

SCIENCE APIS: 30,000 FOOT VIEW

- ▶ Science as a Service
- ▶ Reproducibility
- ▶ Provenance
- ▶ Scalability
- ▶ Advanced access control and collaboration
- ▶ Bring your own data, code, systems, and services
- ▶ Mix and match technologies, services, languages, and env
- ▶ Collaborate across groups and organizations



SCIENCE APIS : KEYBOARD VIEW

1. Conventions
2. Manage data
3. Run code
4. Collaborate meaningfully
5. Integrate anything



SCIENCE APIS : KEYBOARD VIEW

Conventions

- ▶ REST API
- ▶ Sync and Async
- ▶ Event driven
- ▶ JSON in/out
- ▶ Search, response customization, pagination, formatting



SCIENCE APIS : KEYBOARD VIEW

Manage Data

- ▶ Single, consistent interface to access distributed data
- ▶ Managed, tenacious data movement
- ▶ Opinion-free metadata management
- ▶ Full provenance and searchable audit trail.
- ▶ Events, alerts, and notifications
- ▶ Horizontal scaling



SCIENCE APIS : KEYBOARD VIEW

Run code

- ▶ Bring your own code and/or leverage our catalog
- ▶ Run your apps as interactive, batch, or event driven processes
- ▶ Full lifecycle management
- ▶ Full provenance and searchable audit trail
- ▶ Reproducibility as a feature
- ▶ Publish entire experimental runs



SCIENCE APIS : KEYBOARD VIEW

Collaborate meaningfully

- ▶ Secure by default, share as desired
- ▶ Deep link to any resource in the API
- ▶ Generate disposable links to securely share with others
- ▶ Events, webhooks, and web sockets to integrate circa 2016
- ▶ Web standards come standard



SCIENCE APIS : KEYBOARD VIEW

Integrate anywhere

- ▶ Secure by default, share as desired
- ▶ Deep link to any resource in the API
- ▶ Generate disposable links to securely share with others
- ▶ Events, webhooks, and web sockets to integrate circa 2016
- ▶ Web standards come standard





Agave

Tooling

TOOLING: 30,000 FOOT VIEW

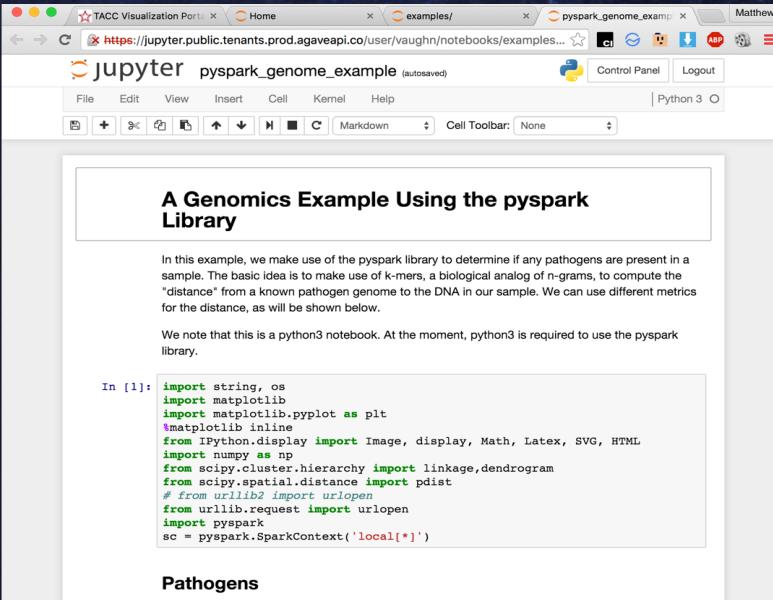
- ▶ Accelerate your development
- ▶ Better utilize the strengths of your dev and tech teams
- ▶ Identify, meet, and mitigate disruptive technology changes before they disrupt you
- ▶ Leverage white label solutions
- ▶ Shrink wrapped integrations with popular technologies
- ▶ `git clone ...` && `docker run...` && PROFIT



TOOLING : KEYBOARD VIEW

- Client SDK: Python, JavaScript, Java, PHP, *Perl*, *R*
- Command Line Interface
- Plugins: AngularJS, Wordpress, Drupal, Tomcat
- Web applications (ToGo)
- Integrated environments (Jupyter Hub)
- Workflow management (End of Day)

TOOLING : SAMPLES & INTEGRATIONS



A Genomics Example Using the pyspark Library

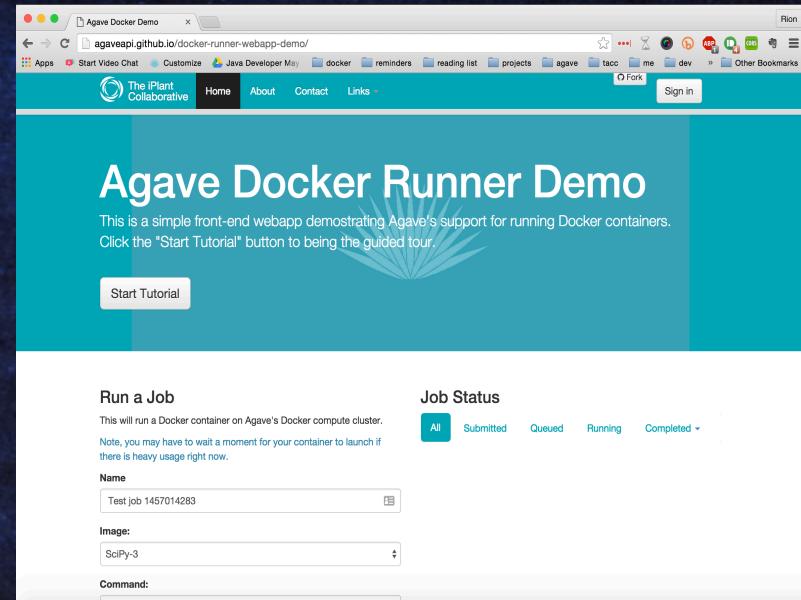
In this example, we make use of the pyspark library to determine if any pathogens are present in a sample. The basic idea is to make use of k-mers, a biological analog of n-grams, to compute the "distance" from a known pathogen genome to the DNA in our sample. We can use different metrics for the distance, as will be shown below.

We note that this is a python3 notebook. At the moment, python3 is required to use the pyspark library.

```
In [1]: import string, os  
import matplotlib  
import matplotlib.pyplot as plt  
%matplotlib inline  
from IPython.display import Image, display, Math, Latex, SVG, HTML  
import numpy as np  
from scipy.cluster.hierarchy import linkage, dendrogram  
from scipy.spatial.distance import pdist  
# from urllib2 import urlopen  
from urllib.request import urlopen  
import pyspark  
sc = pyspark.SparkContext('local[*]')
```

Pathogens

Jupyter integration



Agave Docker Runner Demo

This is a simple front-end webapp demonstrating Agave's support for running Docker containers. Click the "Start Tutorial" button to begin the guided tour.

Start Tutorial

Run a Job

This will run a Docker container on Agave's Docker compute cluster. Note, you may have to wait a moment for your container to launch if there is heavy usage right now.

Name: Test job 1457014283

Image: SciPy-3

Command:

Job Status

All	Submitted	Queued	Running	Completed
1	1	0	0	0

Demos & samples

TOOLING : DOCUMENTATION

Introduction

The Agave Platform (<http://agavepi.co>) is an open source, science-as-a-service API platform for powering your digital lab. Agave allows you to bring together your public, private, and shared high performance computing (HPC), high throughput computing (HTC), Cloud, and Big Data resources under a single, web-friendly REST API.

- Run code
- Manage data
- Collaborate meaningfully
- Integrate anywhere

The Agave documentation site contains documentation, guides, tutorials, and lots of examples to help you build your own digital lab.

If you came here looking for a UI to interact with the platform and kick the tires, please see our reference web application, Agave ToGo, at <http://togo.agavepi.co>.

Conventions

Throughout the documentation you will regularly encounter the following variables. These represent user-specific values that should be replaced when attempting any of the calls using your account. Once you log into this site, these values will be replaced with values appropriate for you to use when copying and pasting the examples on your own.

Variable	Description	Example
\$API_HOST	Base hostname of the API.	public.agavepi.co
\$API_VERSION	Version of the API.	v2

The screenshot shows the Agave API documentation for the `/apps/v2` endpoint. The URL bar at the top shows `http://docs.agaveplatform.org/live-dc`. The main title is "List apps". A "Show samples" button is available. Below the title, there's a "TOKEN" input field containing `ed625654a038fd30fc434c`, and a "GET /apps/v2/" button.

The main content area displays the API schema. It starts with a note: "Get a list of available applications." followed by a "Parameters" section.

Parameters

- public**: `false` boolean
Whether to return only public apps. (Acceptable values are: "true", "false")
- name**:
string The name of the app
- executionSyst...**:
string The execution system of the system
- tags**:
string A free text tag associated with an app
- filter**:
string A comma-separated list of fields in the response objects to return. This allows developers to create their own response objects

RESPONSE SAMPLE

```
{  
  "id": "123456",  
  "name": "My App",  
  "lastModified": "2017-11-04T19:42:58.543Z",  
  "name": "My App",  
  "isPublic": true,  
  "revision": 1,  
  "shortDescription": "string",  
  "version": "string"  
}
```

RESPONSE SCHEMA

Developer guides

Interactive API browser



11/5/17

50



Agave

To Go

AGAVE TOGO

The screenshot shows the Agave ToGo Admin Dashboard. On the left is a sidebar with links for Dashboard, Apps, Projects, Data, Jobs, Lambdas, Webhooks, Systems, Monitors, Community, Feedback, and Support. The main area has a heading "Introducing the new Agave ToGo!" followed by a paragraph about the application's purpose. Below this are four cards: "1349 Jobs" (blue), "12,5TB Data Moved" (red), "549 App Users" (teal), and "+25% Activity Increase" (purple). At the bottom is a "PROJECT ACTIVITY" section with a bar chart showing 3756 total items across categories: USER, COMMENTS, JOBS, DOCS, and KARMA.

<http://togo.agaveplatform.org/app>

The screenshot shows the Agave ToGo MicroSites application. The top navigation bar includes Home, Compute, Data, Remote Sessions, About, Support, and a Search bar. The main content area has a "Welcome" message and a "Agave ToGo MicroSites" section with a "October 31st, 2017" update. Below this is a "PROJECT ACTIVITY" section with a bar chart. At the bottom are "ABOUT", "FOLLOW US ON", and "CONTACTS" sections.

<http://agaveplatform.github.io/microsites/app>

TOOLING : KEYBOARD VIEW

- ▶ Cool kid friendly
 - ▶ Jupyter notebooks
 - ▶ Wordpress, Drupal, Django CMS, AngularJS, Node,...





Agave

App Exchange

APP EXCHANGE: 30,000 FOOT VIEW

- ▶ Library of existing scientific codes
- ▶ Portability across HPC, HTC, Cloud, container, and native environments -- TRUE HYBRID PORTABILITY
- ▶ Write once, run virtually anywhere
- ▶ Social tagging and commentary
- ▶ Benchmarking
- ▶ Based on (~250) existing HPC and HTC apps at TACC



APP EXCHANGE: KEYBOARD VIEW

- ▶ Automated builds to multiple formats
- ▶ Multiple SCM integration
- ▶ Apps built for Docker, Singularity, VM, native runtimes.
- ▶ Automation to run acceptance tests and. Performance benchmarks against known datasets.
- ▶ Consistent execution and monitoring API
- ▶ Form and UI generation
- ▶ Provenance, attribution, fully reproducible runs





OPERATIONAL COSTS

- ▶ License
 - ▶ BSD 2-Clause
- ▶ Services:
 - ▶ monitoring, automation, updates, SLA, backups, security scans, logging, etc
- ▶ Hosting:
 - ▶ db, nosql, cloud storage, app catalog, image registry, public assets, api definitions, service/asset hosting

PEOPLE COSTS

- ▶ Admins:
 - ▶ sql, nosql, tenant, systems, scaling
- ▶ User support:
 - ▶ developer on-boarding
 - ▶ Trainings
 - ▶ docs
 - ▶ workshop design
 - ▶ tech audits
 - ▶ solutions arch.

SUPPORT COSTS

- ▶ Platform
 - ▶ change management
 - ▶ devops
 - ▶ scaling
- ▶ Integration
 - ▶ boutique services
 - ▶ custom integrations
 - ▶ custom IAM
 - ▶ tech evaluations

WHO'S USING THIS STUFF?

SGCI Science Gateways Community Initiative

Science Gateways Catalog

Total Gateways: 226

Search

Log Out (dooley) Add Gateway

FORMAL (16)

- Mathematics (8)
- Chemistry (34)
- Physics (15)
- Earth Sciences (57)
- Space Science (19)
- LIFE (99)
- Biology (88)
- SOCIAL (42)
- Anthropology (2)
- Archaeology (1)
- Criminology (1)
- Geography (2)
- History (1)
- Linguistics (4)
- Pedagogy (1)

Agave ToGo

Agave ToGo v2 is a full-featured web application providing a reference user interface to harnessing the core functionality of the Agave Platform and demonstrate some of the advanced use cases which are...

Brain Life

We are developing an open, online platform to provide a seamless access to cloud computing infrastructure and brain data and data derivatives. This platform is meant to reach out beyond neuroscience. ...

NeuroMorpho.Org

NeuroMorpho.Org is a centrally curated inventory of digitally reconstructed neurons associated with peer-reviewed publications. It contains contributions from over 300 laboratories worldwide and is co...

OSC OnDemand

OSC (Ohio Supercomputer Center) OnDemand is our one-stop-shop for access to our High Performance Computing resources. With OnDemand, you can upload and download files, create, edit, submit, and monitor...

citizenScience.gov

Federal Crowdsourcing and Citizen Science Catalog

The catalog contains information about federal citizen science and crowdsourcing projects. In citizen science, the public participates voluntarily in the scientific process, addressing real-world prob...

SD2E Workbench Learning Center dooley

Synergistic Discovery and Design Environment (SD2E)

ANNOUNCEMENTS: None at present

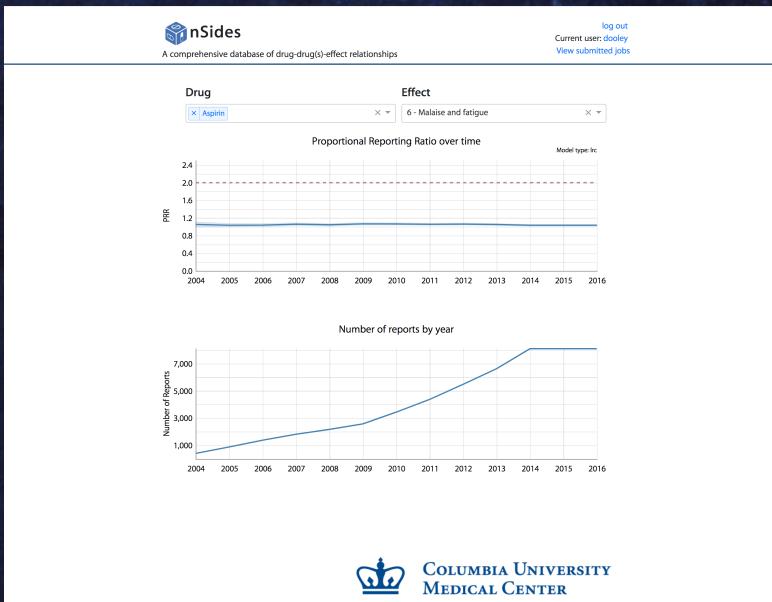
PLATFORM OVERVIEW PROJECT DATA ANALYTICAL ENVIRONMENTS

SD2E.org is the web-based analysis platform for the DARPA SD2 program that enables collaborative analysis, data sharing, and application development backed by national-scale computing, analytics, and storage capacity at the Texas Advanced Computing Center (TACC).

<https://catalog.sciencegateways.org/>

<http://sd2e.org>

WHO'S USING THIS STUFF?



<http://nsides.io>

The screenshot shows the IKE Annotated Repository interface. It features a search bar at the top right, a map view of Hawaii with location markers, and a search results table below. The table lists 23 matching files, each with a download button. The table includes columns for 'Filename', 'Actions', and 'Matching Files: 23'.

<https://ikewai.its.uhawaii.edu>

WHO'S USING THIS STUFF?



The screenshot shows the Arabidopsis Information Portal (AIP) homepage. At the top, there's a navigation bar with links for "About", "Help", "Contact", "Search", and "Log in". Below the navigation is a main menu with "ThaleMine", "Genome Browsers", "Community", "Science Apps", "Downloads", and "Developer Zone". A green banner below the menu says "Explore the Arabidopsis genome". The page features three main sections: "ThaleMine" (with a magnifying glass icon), "JBrowse" (with a magnifying glass icon), and "Science Applications" (with a flask icon). Each section has a brief description and a "More" or "Browse" button.

<http://araport.org>



The screenshot shows the DESIGNSAFE-CI NHERI Community homepage. At the top, there's a logo for "DESIGNSAFE-CI" and "NATIONAL HAZARDS ENGINEERING RESEARCH CENTER". To the right, it says "A CLOUD-BASED ENVIRONMENT FOR RESEARCH IN NATURAL HAZARDS ENGINEERING". Below the logo is a navigation bar with "NHERI Community", "Research Workbench", "Experimental Facilities", "Learning Center", "About", and "Contact". The main content area features a large image of a coastal town under a cloudy sky. Overlaid on the image are the words "NHERI COMMUNITY". Below this, there's a section titled "Find here the latest news and highlights on important natural hazards research discoveries and NHERI program announcements including upcoming meetings, Resources and collaboration tools for the NHERI community will also be provided here." There's also a "News & Features" section with a link to "Register for NHERI Experimental Facilities Workshops 10-10-15".

<https://www.designsafe-ci.org>

WHO'S USING THIS STUFF?

The screenshot shows the CyVerse Discovery Environment interface. On the left, the 'Data' section displays a file tree under 'bigfiles' with several sub-folders like 'bigfiles', 'bigfiles', 'c', and 'example_data'. The 'Analyses' section below lists various bioinformatics tools and their status, such as 'GoSeq v2.0.1', 'Blast 0.0.0', and 'Bowtie2 2.2.5'. Both sections have search and filter options.

The screenshot shows the VDJ Server interface. It features a large circular logo with scientific icons. Below it, four main sections are displayed: 'UPLOAD' (with an up arrow icon), 'ANALYZE' (with a magnifying glass icon), 'PUBLISH' (with a share icon), and 'SHARE' (with a share icon). Each section has a brief description. To the right, there is a 'WELCOME!' message and a login form with fields for 'EMAIL ADDRESS', 'PASSWORD', 'LOGIN', 'Forgot password?', 'Create Account', and 'Send Us Feedback'.

<http://de.cyverse.org>

<https://vdjserver.org>

WHO'S USING THIS STUFF?

FAST TRACK TO DATA ANNOTATION AND BIOINFORMATICS ANALYSIS

Username:
Password:

Log In | Enter As Guest | Forget Password? | Register

D N A
S U B W A Y

Note for:
PLANTS & ANIMALS

Associate & Economic Sequence
Prospect Companies Using TIGER
Reference Sequences Relationships
Next Generation Sequencing

Alignment & Tree Viewer
Bioinformatics & Transfer
Download

This site ties together key bioinformatics tools and databases to assemble gene models, investigate genomes, work with phylogenetic trees and analyze DNA sequences. Roll over the "stations" on the subway map to find out more about the analysis maps. Analyze your own data or sample data provided. To start a project, select a station, log into your account, register and login to be able to save and share your work.

<http://cyverse.org>

► DNA Subway Training ► DNA Revealing DNA
● Background ● Manual ● Tutorials
► About ► Credits ► Resources ► Contact Us
Plant Collaborators

Cold Spring Harbor Laboratory

BioExtract Server
data access, analysis, storage, and workflow creation

Send us Feedback
Contact User: guest
[Sign In | Register | Why Register?]

Query Extracts Tools Workflows Groups Help

Demo Workflow

Available Data Sources. Select one or more data sources to query:

- + All
- + Miscellaneous
- + Nucleotide Sequences
- + Protein Sequences
- + Viridiplantae
- + Viridiplantae Protein

Fetch Sequence(s)
If you know the GI accession number(s) of the sequence(s) you want to retrieve, the Fetch Sequence Records tool will retrieve records and display on the Extracts page.

What's New

[View](#) [Facebook](#) [Twitter](#)

Query Form. Select a search field and enter a search term.
Press Add Search Line to combine search terms with AND, OR, and NOT. Query examples:
[Add Search Line](#)

Search Field: (all types)

Search Term(s):

Current Query:

Version: 2.4.6 (Unreleased)

About BioExtract harnesses the power of online bioinformatics tools for creating and customizing workflows. Users can query online sequence data, analyze it using an informatics tools (web service and desktop), create and share custom workflows for repeated analysis, and save the resulting data and workflows. This work was initially supported by NSF grants 0649803, 0704113, and 0835488. Current work is being supported by NIH 2008-1135488.

<http://dnasubway.org>

<https://www.bioextract.org>

WHO'S USING THIS STUFF?

The screenshot shows the homepage of the iReceptor website. At the top is a green header bar with the iReceptor logo, which features a stylized Y-shaped icon made of green dots. Below the header is a navigation menu with links for About, News, Architecture, and Contact. The main content area has a white background. On the left, there's a section titled "What is iReceptor?" containing a brief description of the project. To the right, there's a "LATEST NEWS" section with three items: "CANARIE Workshop Oct 30-21, 2014 10/20/2014 - 08:59", "Antibody & T-cell Receptor Data Integration Planning Meeting 09/10/2014 - 15:07", and "iReceptor CANARIE announcement 06/23/2014 - 11:31". At the bottom right, it says "iReceptor is live!! 06/22/2014 - 08:48".

<https://ireceptorgw.irmacs.sfu.ca/>

The screenshot shows the homepage of the iMicrobe website. It features a banner with the text "iMicrobe Enables Science" and "Search, analyze, compare, and visualize your data!". Below the banner are two orange buttons: "GET STARTED" and "LOGIN". The main content area has a white background. It includes a section titled "Use cases" with four items: "Browse" (described as a "window shop" for datasets by ecosystem, project, principal investigator, and more), "Search" (described as how-in-on datasets based on environmental factors using our powerful search tool), "Discover" (described as using our analytics to find datasets like yours based on sequence composition and shared functional or taxonomic annotations), and "Recommendations" (described as joining the iMicrobe community and getting personalized recommendations on recent articles in microbial ecology). There are also sections for "Learn" and "Developers".

<https://www.imicrobe.us>



THANK YOU!

FOLLOW US

@agaveapi

<https://agaveapi.co>