



# SCIENCE-AS-A-SERVICE WITH THE AGAVE PLATFORM

Rion Dooley @deardooley  
[dooley@tacc.utexas.edu](mailto: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

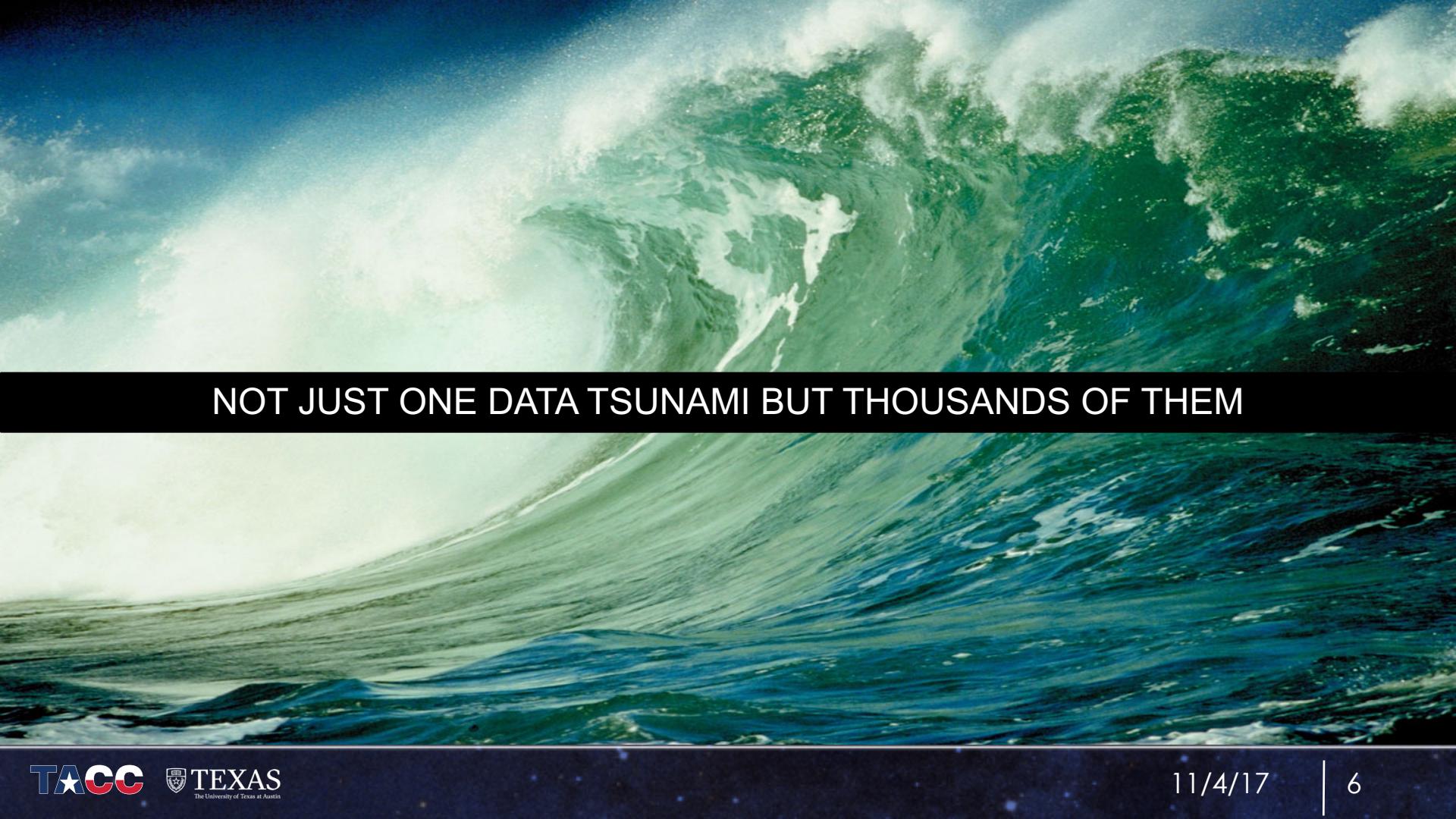


*But a curious thing happened while HPC was running its victory lap...*

*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**

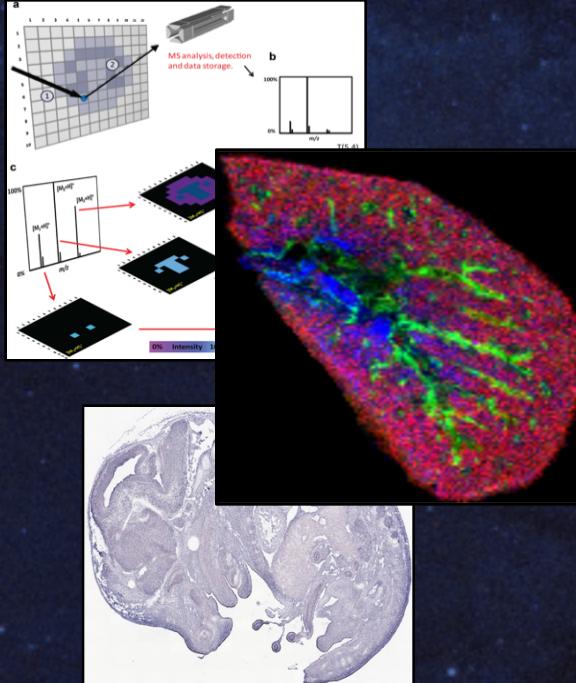
- Mostly experimentalists
- Strict data sharing & access



**Roshan**

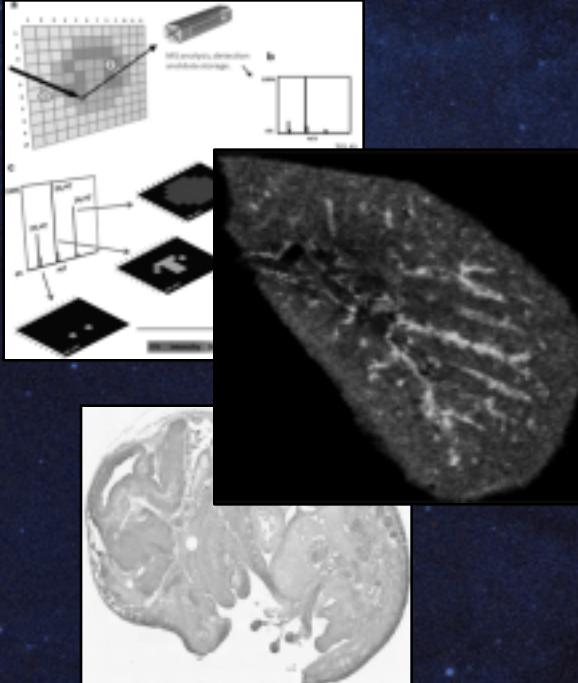
- Computationally experienced
- Focused on interpretation

# 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)



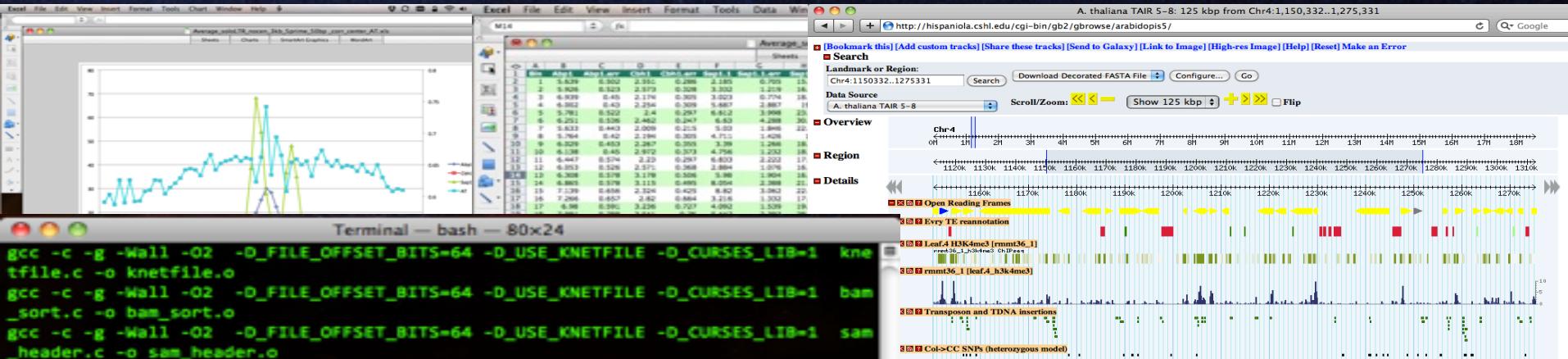
- ▶ Data lifecycle management
    - ▶ Fine-grained permission management
    - ▶ Discoverability
    - ▶ Version control
  - ▶ Domesticating promising new analysis codes based on often immature technology
  - ▶ Doing reproducible computational science
  - ▶ Adopting efficient analytical methods

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



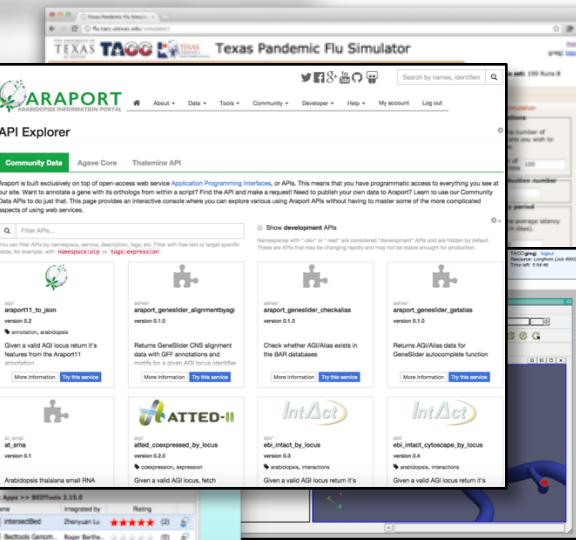
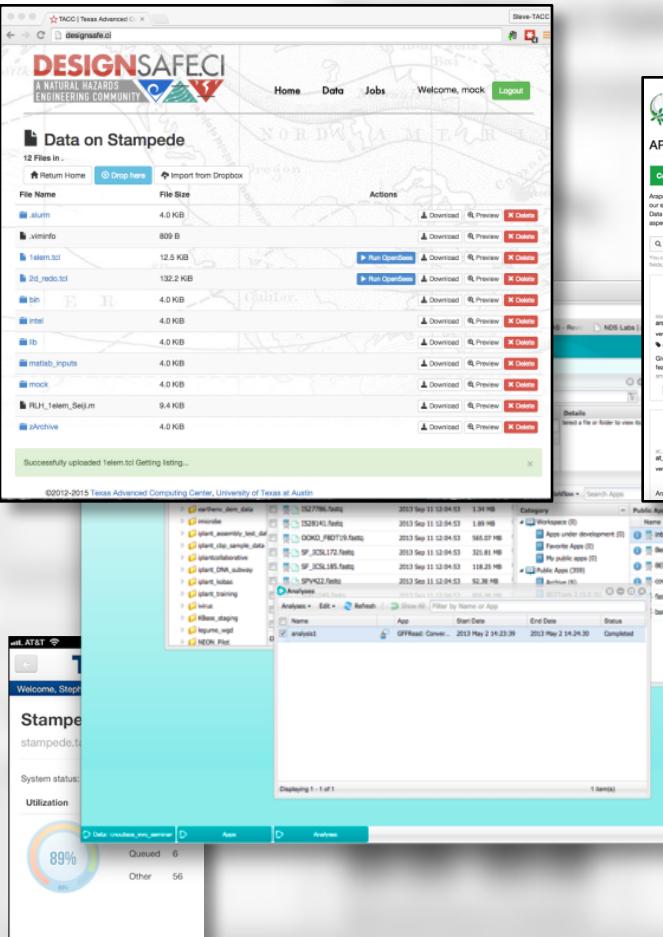
BUILD A MASSIVE STORAGE CLOUD NEXT TO INNOVATIVE, POWERFUL, USABLE COMPUTERS AT THE END OF FAST INTERNET PIPES





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





# 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

The screenshot shows a desktop environment with two windows open. On the left is an RStudio session window titled 'Console ~/testing/'. It displays the R command-line interface with the following output:

```
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
You are welcome to redistribute it under certain
Type 'license()' or 'licence()' for distribution
Natural language support but running in an En
```

Below this, there is more R documentation and a command line prompt.

On the right is a Jupyter notebook window titled 'jupyter pyspark\_genome\_example (autosaved)'. The title bar includes 'Control Panel' and 'Logout' buttons. The notebook interface has tabs for File, Edit, View, Insert, Cell, Kernel, and Help, and dropdown menus for Python 3, Cell Toolbar, and Markdown.

## 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[*]')
```

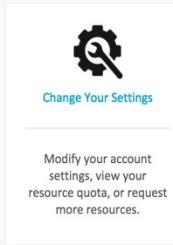
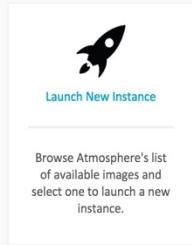
## Pathogens

## Hosted SaaS

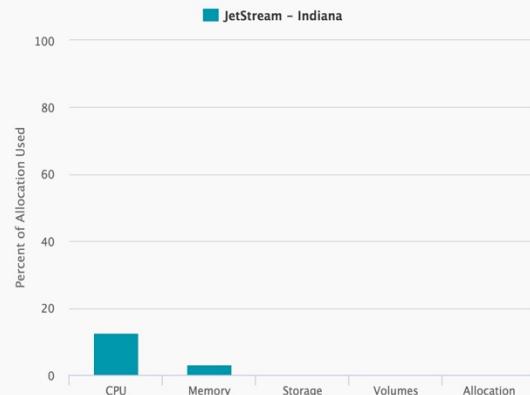
- JupyterHub notebooks
- Rstudio
- Web-based VNC

Also, backed by world-class computing and data capacity

## Getting Started



## Resources in Use [Need more?](#)



1 Instances

active - deploy\_error



0 Volumes

## Community Activity

**edwintest3** created an image Nov 16, 2015 02:31 am [MAKER-P 2.28 with CCTools 5](#)

**edwintest3** created an image Nov 16, 2015 02:31 am [TSW Workshop Williams 1.2](#)

**atmodadmin** created an image Oct 23, 2015 12:06 am [Trusty Tahr \(x64\)](#)

**atmodadmin** created an image Oct 23, 2015 12:06 am [cirros-0.3.4-x86\\_64](#)

**atmodadmin** created an image Oct 23, 2015 12:06 am [CentOS-7-x86\\_64-GenericCloud-20150628\\_01](#)

**atmodadmin** created an image Oct 23, 2015 12:06 am [CentOS-6-x86\\_64-GenericCloud-1508](#)

**atmodadmin** created an image Oct 23, 2015 12:06 am [CentOS-7-x86\\_64-GenericCloud-1508](#)

## 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

The logo for the Agave Platform. It features the word "Agave" in a large, white, sans-serif font. The letter "A" is unique, containing a stylized graphic of several thin, radiating lines that resemble a sunburst or a cluster of agave plant leaves. Below "Agave", the word "Platform" is written in a smaller, white, sans-serif font.

# Agave

Platform

<https://github.com/agaveplatform/nectar-2017-workshop>

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 KICKING THE TIRES LOOKS LIKE

## MONTHLY USAGE

30 compute sites  
100 clients  
1,000 apps  
6,000 simulations  
1,000,000 transfers  
2,500,000 GB data moved

## AVG. MONTHLY GROWTH

45 new systems  
30 new clients  
50 new apps  
2000 docs shared  
100,000 new transfers  
250,000 GB data moved

### WORLDWIDE PLATFORM USAGE

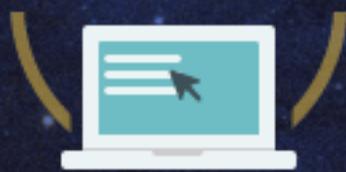


# WHAT DOES IT DO?

MANAGE  
DATA



RUN  
CODE



COLLABORATE  
ANYWHERE



CONNECT  
ANYTHING



# WHAT CAN AGAVE DO FOR ME?





The logo for the Agave Platform. It features the word "Agave" in a large, white, sans-serif font. The letter "A" is unique, containing a stylized graphic of several thin, white lines radiating from a central point, resembling a fan or a starburst. Below "Agave", the word "Platform" is written in a smaller, white, sans-serif font.

# 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 - Full implementation
- ▶ Federated login
- ▶ SSO
- ▶ MFA
- ▶ Custom flows



# PLATFORM : KEYBOARD VIEW

## API Management

- ▶ SSL termination
- ▶ Gateway proxy
- ▶ Load balancing
- ▶ JWT generation
- ▶ Lifecycle management
- ▶ Swagger 2 API support
- ▶ Metrics
- ▶ 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 screenshot of a Jupyter notebook interface. The title bar shows the URL [https://jupyter.public.tenants.prod.agaveapi.co/user/vaughn/notebooks/examples/pyspark\\_genome\\_example.ipynb](https://jupyter.public.tenants.prod.agaveapi.co/user/vaughn/notebooks/examples/pyspark_genome_example.ipynb). The notebook content is titled "A Genomics Example Using the pyspark Library". It contains text explaining the use of the pyspark library to determine if any pathogens are present in a sample, mentioning k-mers and n-grams. Below this is a code cell labeled "In [1]" containing Python code to import various libraries like string, os, matplotlib, and pyspark. At the bottom, a section titled "Pathogens" is visible.

```
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[*]')
```

Jupyter integration

A screenshot of a web application titled "Agave Docker Runner Demo". The page has a teal header with the title "Agave Docker Runner Demo" and a sub-header "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." Below the header is a "Start Tutorial" button. The main area is titled "Run a Job" and contains instructions: "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." It includes fields for "Name" (set to "Test job 1457014283"), "Image" (set to "SciPy-3"), and "Command". To the right is a "Job Status" section with tabs for "All", "Submitted", "Queued", "Running", and "Completed".

Demos & samples

# TOOLING : DOCUMENTATION

# Introduction

The Agave Platform (<http://agaveapi.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.agaveapi.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
<code>\$API_HOST</code>	Base hostname of the API.	<code>public.agaveapi.co</code>
<code>\$API_VERSION</code>	Version of the API.	<code>v2</code>

**URL**

<http://docs.agaveplatform.org/live-dc>

**TOKEN**

ed62565f4a038b330f434ac4 →

**GET /apps/v2/**

Get a list of available applications.

**Parameters**

Clients

Apps

**List Apps**

Register And Update New Application...

Deletes An Application.

Get Details Of An Application By Its...

Update An Application.

Edit An Application.

Deletes All Permissions On An Application...

Get The Permission For This Application...

Grant A User Permission For An Application...

Deletes All Permissions For The Given Application...

Get A Specific User Permission For An Application...

Add Or Update A User Permission For An Application...

Get A Submission Form For The Application...

List The Event History Of This App

Files

Jobs

Meta

Monitors

Notifications

Postits

Profiles

**List apps**

Show samples >

**RESPONSE SAMPLE**

```
[{"id": "string", "executionSystem": "string", "lastModified": "2017-11-04T19:02:50.543Z", "name": "string", "isPublic": true, "revision": 1, "shortDescription": "string", "version": "string"}]
```

**RESPONSE SCHEMA**

public false boolean Whether to return only public apps. (Acceptable values are: "true", "false")

name string The name of the app

executionSystem 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

# Developer guides

# Interactive API browser



# Agave

To Go

# AGAVE TOGO

The screenshot shows the Agave ToGo Admin Dashboard. On the left, a sidebar lists navigation items: Dashboard, Apps, Projects, Data, Jobs, Lambdas, Webhooks, Systems, Monitors, Community, Feedback, and Support. The main area features a large "Introducing the new Agave ToGo!" section with four cards: "1349 Jobs" (blue), "12,5TB Data Moved" (red), "549 App Users" (teal), and "+25% Activity Increase" (purple). Below this are two charts: "PROJECT ACTIVITY" (bar chart) and "ACTIVITY FEEDS" (list of items: Total 3756, New 1, USER, COMMENTS, JOBS, DOCS, KARMA).

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

The screenshot shows the Agave ToGo MicroSites application homepage. The header includes the Agave logo and navigation links: Home, Compute, Data, Remote Sessions, About, Support, and Search. The main content area has a "Welcome" message and a "MicroSites" section. The footer contains "ABOUT", "FOLLOW US ON", and "CONTACTS" sections, along with a copyright notice: "2016 © Texas Advanced Computing Center".

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

# TOOLING : KEYBOARD VIEW

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





The logo for Agave App Exchange features the word "Agave" in a large, white, sans-serif font. The letter "A" is unique, containing a stylized graphic of several thin, radiating lines that resemble a sunburst or a cluster of leaves. Below "Agave", the words "App Exchange" are written in a smaller, white, sans-serif font.

# 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 (11)
- 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

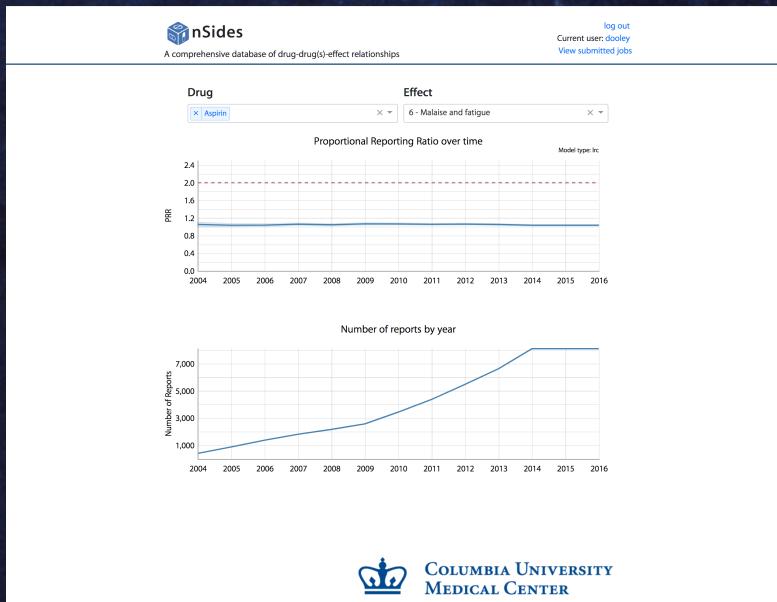
PLATFORM OVERVIEW PROJECT DATA ANALYTICAL ENVIRONMENTS

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. On the left, a sidebar includes 'Data', 'Data Descriptors', 'Share Links', 'Search' (selected), 'Text Search', 'Faceted Search', 'Map Search', and 'Help'. The main area has a search bar 'Search Ike Annotated Repository Files' and a map of Hawaii with several location markers. Below the map is a section titled 'Matching Files: 23' with a table of results:

Filename	Actions
/new_data/DS_Store	<a href="#">Download</a>
/new_data/01-51-07	<a href="#">Download</a>
/new_data/3-1642-001-.pdf	<a href="#">Download</a>
/new_data/3-1642-001.pdf	<a href="#">Download</a>
/new_data/3-1642-001.pdf-cp7	<a href="#">Download</a>
/new_data/3-1646-001.pdf	<a href="#">Download</a>

<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 green header bar with the text "Explore the Arabidopsis genome". The main content area features three cards: "ThaleMine" (with a magnifying glass icon), "JBrowse" (with a magnifying glass icon), and "Science Applications" (with a flask icon). Each card has a brief description and a blue "More" or "Browse" button.

<http://araport.org>



The screenshot shows the DESIGNSAFE-CI NHERI Community homepage. At the top, there's a navigation bar with links for NHERI Community, Research Workbench, Experimental Facilities, Learning Center, About, and Contact. The main content area features a large banner with the text "DESIGNSAFE-CI" and "A CLOUD-BASED ENVIRONMENT FOR RESEARCH IN NATURAL HAZARDS ENGINEERING". Below the banner is a section titled "NHERI COMMUNITY" with a description of the latest news and highlights. To the right is a photograph of a coastal town under construction or repair after a disaster.

<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 shows a list of completed analyses, including 'GoSeq v2.0\_analysis1' and 'MergeGP 0.0.1\_analysis1', along with their owners, apps used, start and end dates, and status.

Data bigfiles Apps Analyses

The screenshot shows the VDJ Server interface. It features a search bar at the top and a list of software tools categorized by type. The 'High-Performance Computing' section includes tools like 'bypass 2.0', 'Bismark (genome preparation) 0.14.4', 'Bismark 0.14.4', 'Bisulfit 0.1.2', 'Blast 0.0.0', and 'Bowtie2 2.2.5'. Each entry includes a thumbnail, name, version, owner, and status.

The screenshot shows the VDJ Server landing page. It features a large 'WELCOME!' message and a circular graphic illustrating molecular biology concepts. Below are four main sections: 'UPLOAD' (with an icon of an upward arrow), 'ANALYZE' (with an icon of a magnifying glass over a DNA helix), 'PUBLISH' (with an icon of a document with a checkmark), and 'SHARE' (with an icon of a person sharing a document). Each section has a brief description.

<http://de.cyverse.org>

<https://vdjserver.org>

# WHO'S USING THIS STUFF?

The DNA Subway interface features a "FAST TRACK TO DATA ANNOTATION AND BIOLOGICAL ANALYSIS" header. It includes a login form with fields for Username and Password, and links for Log In, Enter As Guest, Forget Password?, Register, and a "DNA SUBWAY" logo. A "Note for PLANTS & ANIMALS" section is shown on the right. The main area displays four parallel "subway" tracks representing different bioinformatics workflows:

- Associate to Economic Sequence**: Red track.
- Predict Consensus Using TIGRFAM**: Orange track.
- Sequence Relationships**: Blue track.
- Next Generation Sequencing**: Green track.

Each track has associated nodes and arrows indicating data flow. A central "Alignment & Tree Viewer" node is connected to all tracks. A "Browsers & Transfer" node is at the end of the blue and green tracks. A "Click to watch" button is located near the green track. A descriptive text block at the bottom explains the site's purpose: "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>

The BioExtract Server interface has a header with "BioExtract Server" and "data access, analysis, storage, and workflow creation". It includes links for Query, Extracts, Tools, Workflows, Groups, and Help, along with a "Send us Feedback" link and a "Guest" user note. A "Demos Workflow" link is also present.

The main content area is titled "Available Data Sources". It lists several options with checkboxes:  
+ All  
+ Miscellaneous  
+ Nucleotide Sequences  
+ Protein Sequences  
+ Viridiplantae  
+ Viridiplantae Protein

A "Fetch Sequence(s)" section on the right provides instructions for retrieving sequences by accession number. Below it is a "What's New" section with social media links.

The "Query Form" section contains a "Search Field" dropdown set to "All Type", a "Search Term(s)" input field, and a "Current Query" section with "Submit Query" and "Clear" buttons.

At the bottom, a "Version: 2.4.6 (Unreleased)" note and a "About Us" section with funding information are displayed.

<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 antibody molecule. Below the header is a navigation menu with links for About, News, Architecture, and Contact. The main content area has a large heading "What is iReceptor?" followed by a detailed paragraph about the project's goal of mining "Next Generation" sequence data from immune responses to design vaccines and therapeutic antibodies. Below this is another paragraph about the platform's purpose and components. To the right, there's a sidebar titled "LATEST NEWS" with three entries: "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 of the page is a small note: "iReceptor is live!! 06/22/2014 - 08:48".

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

The screenshot shows the homepage of the iMicrobe website. The header includes the iMicrobe logo, a search bar, and links for Browse, Tools, Download, and Login. A note at the top states, "Note: This is a new beta version and not currently under development. Please let us know of any issues or suggestions!" The main banner features a blue-toned image of various microorganisms and the text "iMicrobe Enables Science" with a subtext "Search, analyze, compare, and visualize your data!". Below the banner are two orange buttons: "GET STARTED" and "LOGIN". A section titled "Use cases" lists five categories with icons: "Browse" (a magnifying glass), "Search" (a magnifying glass), "Discover" (a lightbulb), "Recommendations" (a person icon), and "Learn" (a book icon). Each category has a brief description and a "Read more" link.

<https://www.imicrobe.us>



THANK YOU!

FOLLOW US

@agaveapi

<https://agaveapi.co>