

*Connecting people and resources to
accelerate discovery by empowering the
science gateway community*



Science Gateways Community Institute

*Suresh Marru
Indiana University*

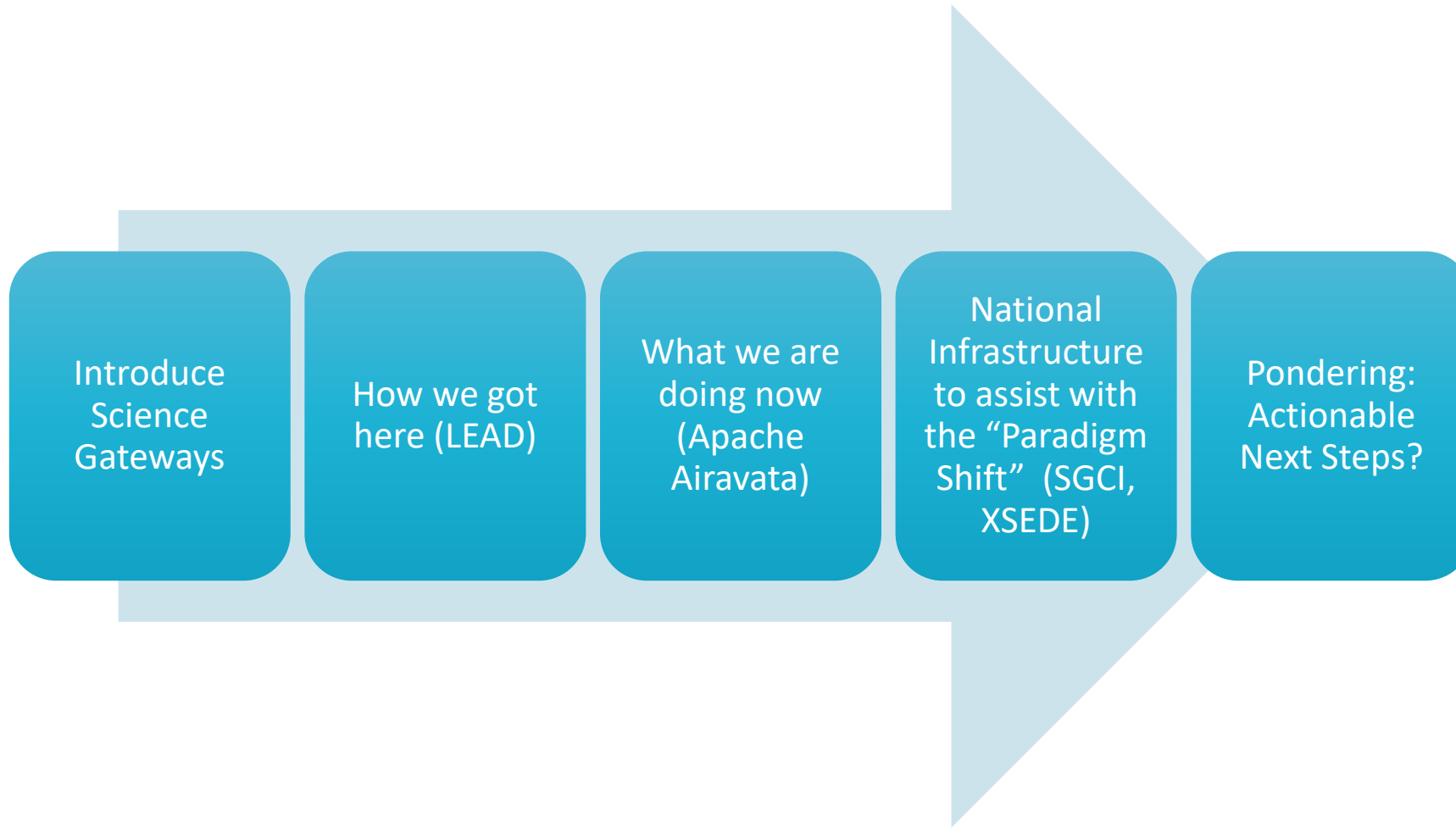


Award Number
ACI-1547611

About me

- Deputy Director for the Science Gateways Research Center
- Nominated Member of the Apache Software Foundation
- Co-PI on several NSF awards
- Co-Instructor, Science Gateways Architectures Course
- The Apache Software Foundation Vice President for Apache Airavata

Aims for next 30 mins



What is a science gateway?

science gateway /sī' əns gāt' wā/ *n.*

1. an online community space for science and engineering research and education.
2. a Web-based resource for accessing data, software, computing services, and equipment specific to the needs of a science or engineering discipline.



Computational Science and Engineering Challenges

- What is a cluster/cloud and how do I use it?
- What clusters/cloud are available to me?
- How do I use this particular machine?
- How do I get my data on and off?
- How did I get that result?
- Where is that result?
- Can you share that result with me?

Problems

- My applications are taking too long to run on my desktop.
- I know I should run my applications on supercomputer, but it is too confusing, and the person who knows how to do that is too busy.
- You mean there are other supercomputers or computing clouds out there that I could use besides my team's cluster?
- That supercomputer is really different from what I know how to use, so I'll not bother.
- *[Data input for my application on Supercomputer A is different Supercomputer B.*
 - *Workflow problem: connect A to B]*

Science Gateways Solve These Problems

Technology Adoption Choices

```
import sys
conf = json.load(open('tasconf.json','r'))
tas_session = requests.Session()
tas_project_list = [x['chargeCode'] for x in tas_session.get(conf['api_url'] + '/projects/resource/Jetstream', auth=('tas-jetstream', conf['tas_pass'])).json()['result']]
auth = v3.Password(auth_url=conf['as_auth_url'], user_id=conf['as_user_id'], password=conf['as_password'], project_id=conf['as_project_id'])
sess = session.Session(auth=auth)
keystone = client.Client(session=sess)
connection = pymysql.connect(host=conf['mysql_host'], user=conf['mysql_user'], passwd=conf['mysql_pass'], db='ceilometer')
cursor = connection.cursor()
cursor.execute('select id,generated from event where event_type_id=28 and generated > %f order by generated' % conf['last_generated'])
su_table = {'m.tiny': 1, 'm.small': 2, 'm.medium': 6, 'm.large': 18, 'm.xlarge': 24, 'm.xxlarge': 44}
project_cache = {}
user_cache = {}
for (item,generated) in cursor.fetchall():
    try:
        cursor.execute('select trait_text.value from trait_text where event_id=%d and trait_text.key="state"' % item)
        if cursor.fetchone()[0] == 'active':
            cursor.execute('select trait_text.value from trait_text where event_id=%d and trait_text.key="project_id"' % item)
            project_id = cursor.fetchone()[0]
            cursor.execute('select trait_text.value from trait_text where event_id=%d and trait_text.key="user_id"' % item)
            user_id = cursor.fetchone()[0]
            cursor.execute('select trait_text.value from trait_text where event_id=%d and trait_text.key="host"' % item)
            host = cursor.fetchone()[0]
            cursor.execute('select trait_text.value from trait_text where event_id=%d and trait_text.key="instance_id"' % item)
            instance_id = cursor.fetchone()[0]
            cursor.execute('select trait_text.value from trait_text where event_id=%d and trait_text.key="instance_type"' % item)
            instance_type = cursor.fetchone()[0]
            cursor.execute('select trait_datetime.value from trait_datetime where event_id=%d and trait_datetime.key="audit_period_beginning"' % item)
            audit_period_beginning = datetime.datetime.fromtimestamp(cursor.fetchone()[0])
            cursor.execute('select trait_datetime.value from trait_datetime where event_id=%d and trait_datetime.key="audit_period_ending"' % item)
            audit_period_ending = datetime.datetime.fromtimestamp(cursor.fetchone()[0])
            cursor.execute('select trait_datetime.value from trait_datetime where event_id=%d and trait_datetime.key="launched_at"' % item)
            launched_at = datetime.datetime.fromtimestamp(cursor.fetchone()[0])
            su = su_table[instance_type] * (audit_period_ending - audit_period_beginning).total_seconds()/3600
            if project_id not in project_cache:
                try:
                    project_cache[project_id] = keystone.projects.get(project_id).name
                except:
                    project_cache[project_id] = project_id
            if project_cache[project_id] in tas_project_list:
                if user_id not in user_cache:
                    user_cache[user_id] = keystone.users.get(user_id).name
                d = {'endUTC': audit_period_ending.strftime('%Y-%m-%dT%H:%M:%S'), 'project': project_cache[project_id], 'queueName': host, 'queueUTC': launched_at.strftime('%Y-%m-%dT%H:%M:%S'),
                    'resource': 'Jetstream', 'schedulerId': instance_id + '-' + item, 'startUTC': audit_period_beginning.strftime('%Y-%m-%dT%H:%M:%S'), 'sus': su, 'username': user_c
                    'cache[user_id], 'cpus': su_table[instance_type]}
                print d
```

Nanocad Editor

Summary of Nanocad Commands:

Rotate: drag gray space	Translate: Shift-drag gray space	Zoom: Ctrl-drag gray space
Move Atom: drag atom	Add Atom: Shift-click gray space	Delete Atom: Shift-click atom
Add Bond: Shift-drag atom to atom	Delete Bond: Ctrl-drag atom to atom	Select Atom: Alt-click atom
Add double bond: Shift-drag between bonded atoms		Select Group: Ctrl-Alt-click atom

About

Import Structure

Atom Database My Files Function-Group Ion **Molecule**

Change current element to: **H** Select

Biology Inorganic Organic

Aminoacid

- alanine
- arginine
- asparagine
- aspartate
- cysteine
- glutamate
- glutamine
- glycine
- histidine
- isoleucine
- leucine
- lysine
- methionine
- phenylalanine
- proline
- serine
- threonine
- tryptophan
- tyrosine
- valine

Group Geometry Forces Help Structure Clear Undo

Get Potential --Minimize-- --Force Field-- --Input/Output Menu--

Show atom information here ...

What is a Science Gateway?

- Science gateways are Web and desktop interfaces to high performance computing clusters, computing clouds.
- Science gateways encode expertise
 - Running specific scientific application
 - Running jobs on diverse, nonlocal machines
 - Moving data to and from world-wide resources
- Science gateways enable sharing of results
- Science gateways make results recoverable and reproducible

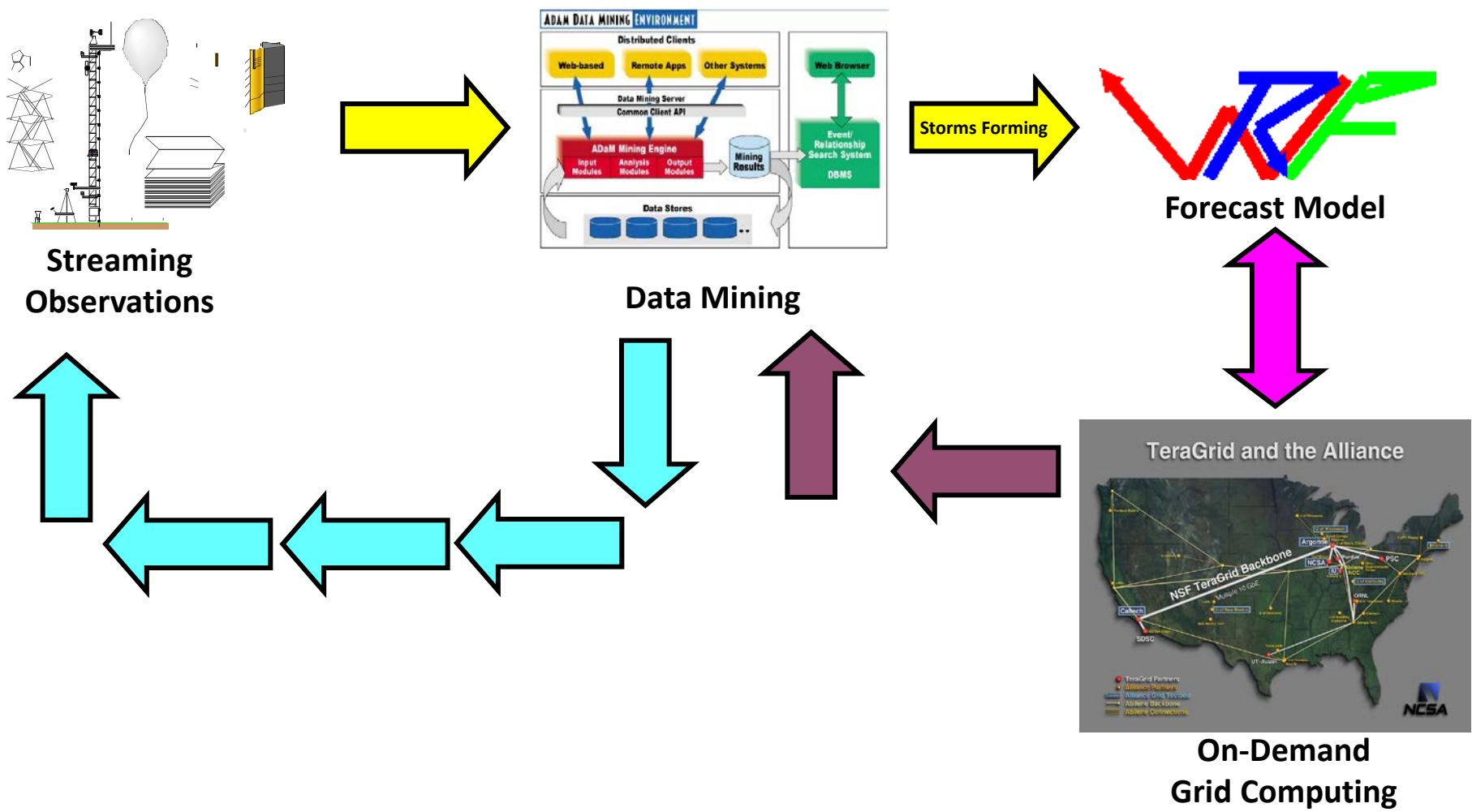
Science Gateways Are So Popular
that We Started a Center
And then “Apache Airavata is
Software that we developed to
build science gateways”



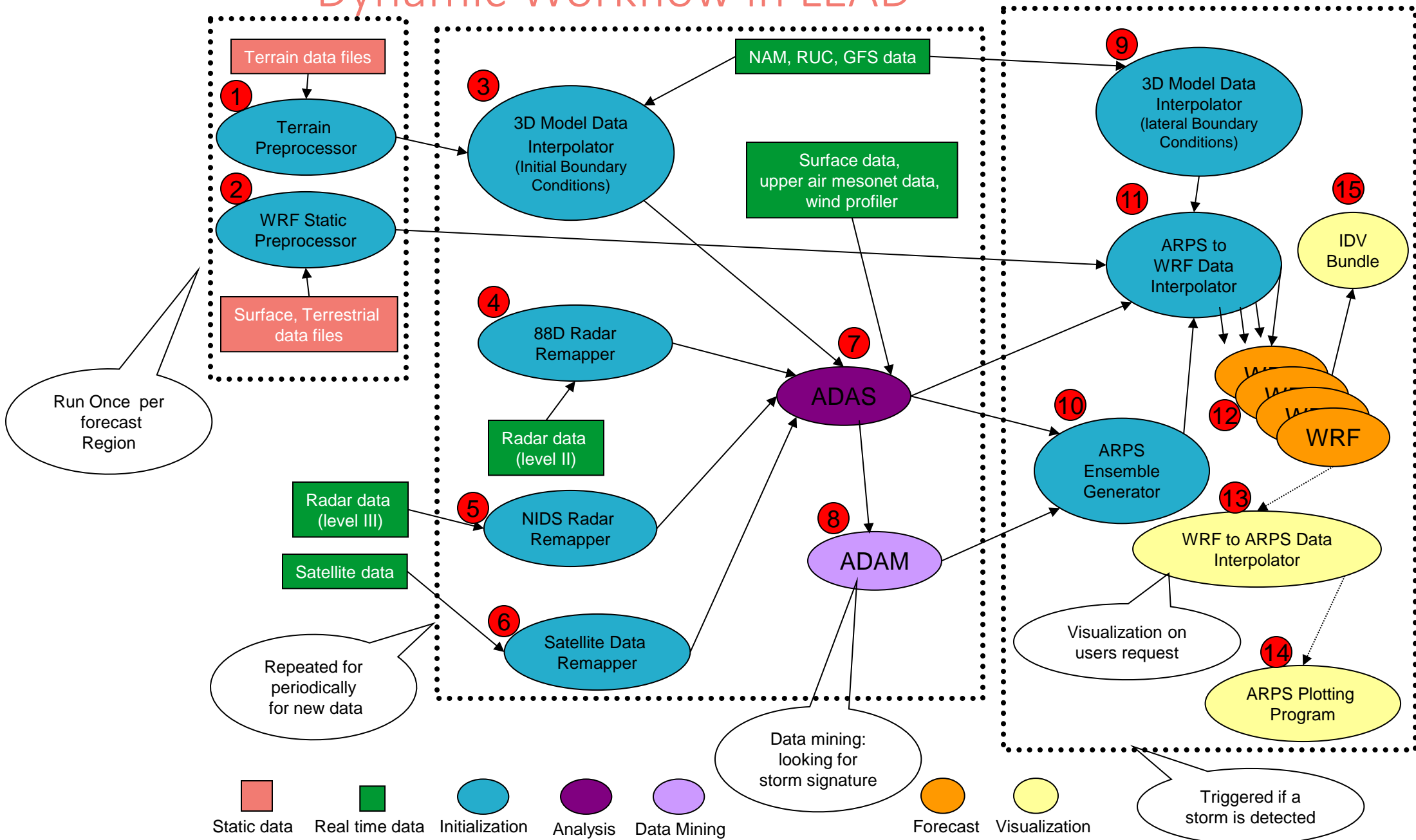
LINKED
ENVIRONMENTS
FOR ATMOSPHERIC
DISCOVERY



Example: Adapting Weather Prediction to Observational Sources Using Dynamic Adaptivity



Dynamic Workflow in LEAD



LEADPORTAL
LINKED ENVIRONMENTS FOR ATMOSPHERIC DISCOVERY

SPONSORED BY THE NATIONAL SCIENCE FOUNDATION

HOME MY WORKSPACE ABOUT LEAD DATA SEARCH RESOURCES HELP

Welcome

WELCOME TO THE LEAD PORTAL

Linked Environments for Atmospheric Discovery (LEAD) makes meteorological data, forecast models, and analysis and visualization tools available to anyone who wants to interactively explore the weather as it evolves. The LEAD Portal brings together all the necessary resources in a convenient access point ... [read more](#)

FEATURES FOR ANYONE INTERESTED IN THE WEATHER

Research affiliations GET FEATURES

Educational institutions, or middle schools GET FEATURES

Students high school levels GET FEATURES

Visitors Newcomers to the curious GET FEATURES

POPULAR TOOLS

Visualize Weather Data
[Integrated Data Viewer](#) | MORE >

Create a Forecast or Analysis
[Experiment Builder](#) | MORE >

Access Weather Data
[Geographic Region Search](#) | MORE >

QUICK LINKS

- Live Weather
- LEAD Grid
- LEAD Blog
- Glossary
- Website Help
- Frequently Asked Questions
- LEAD (password protected)

THE LEAD TEAM

Colorado State University HOWARD UNIVERSITY INDIANA UNIVERSITY MILLERSVILLE UAH UCAP ILLINOIS OKLAHOMA

ANALYZE & PREDICT

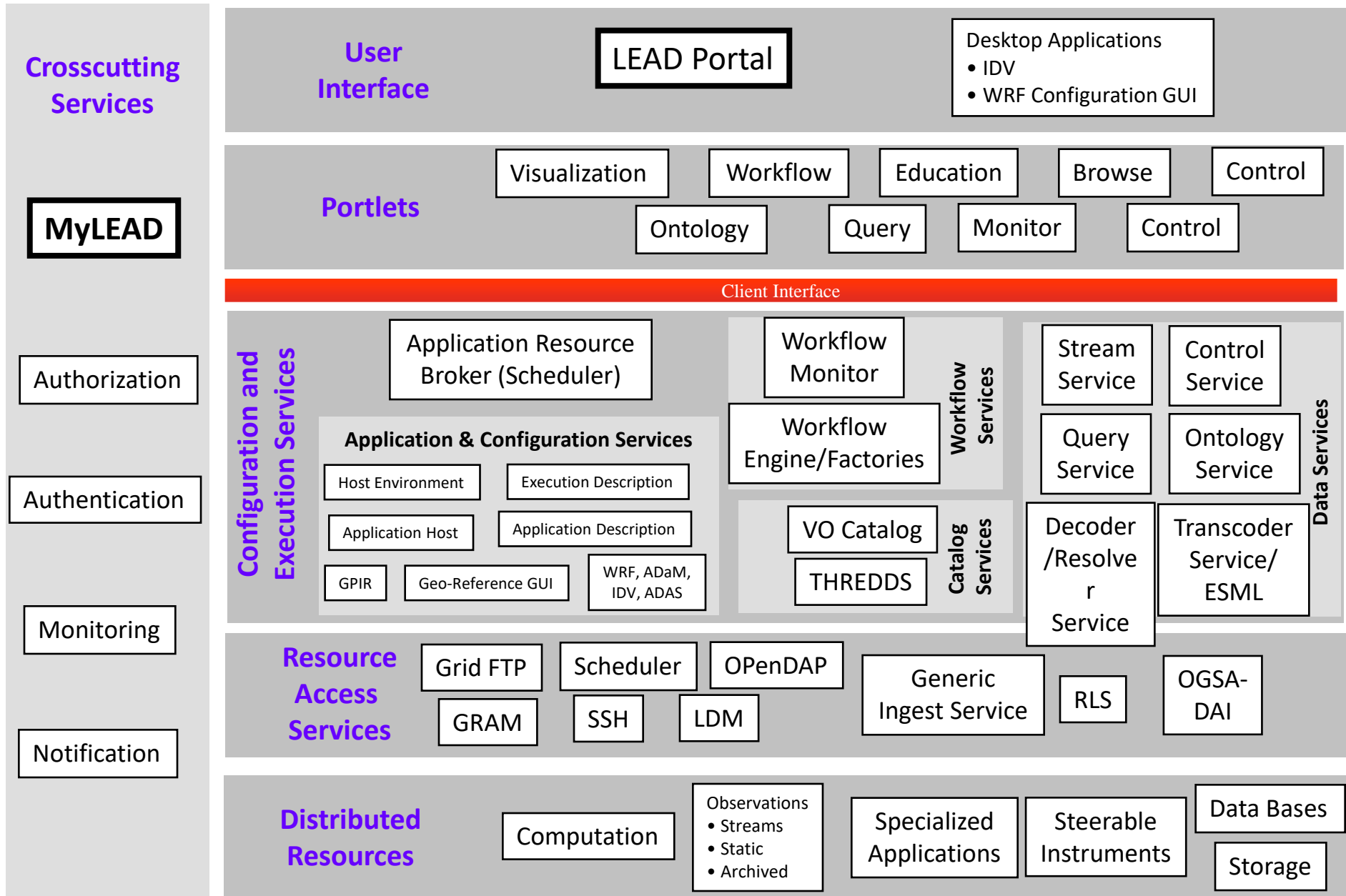
DISCOVER & VISUALIZE

RESEARCH & REPRODUCIBILITY

EDUCATION & OUTREACH

COPYRIGHT © 2006 LINKED ENVIRONMENTS FOR ATMOSPHERIC DISCOVERY. ALL RIGHTS RESERVED. CONTACT US | LOGOUT

LEAD Architecture

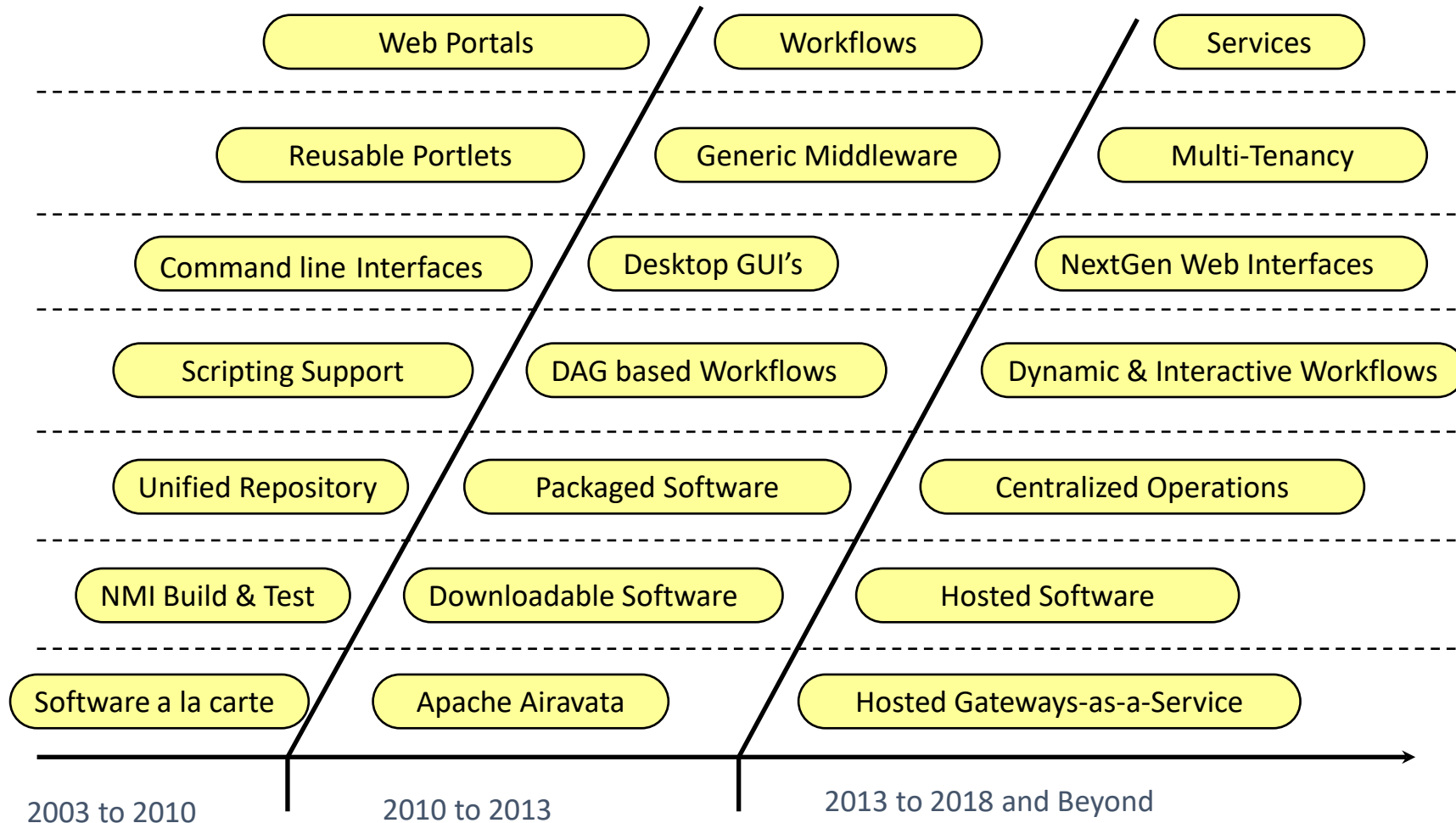


Excerpt from LEAD Final Report to NSF

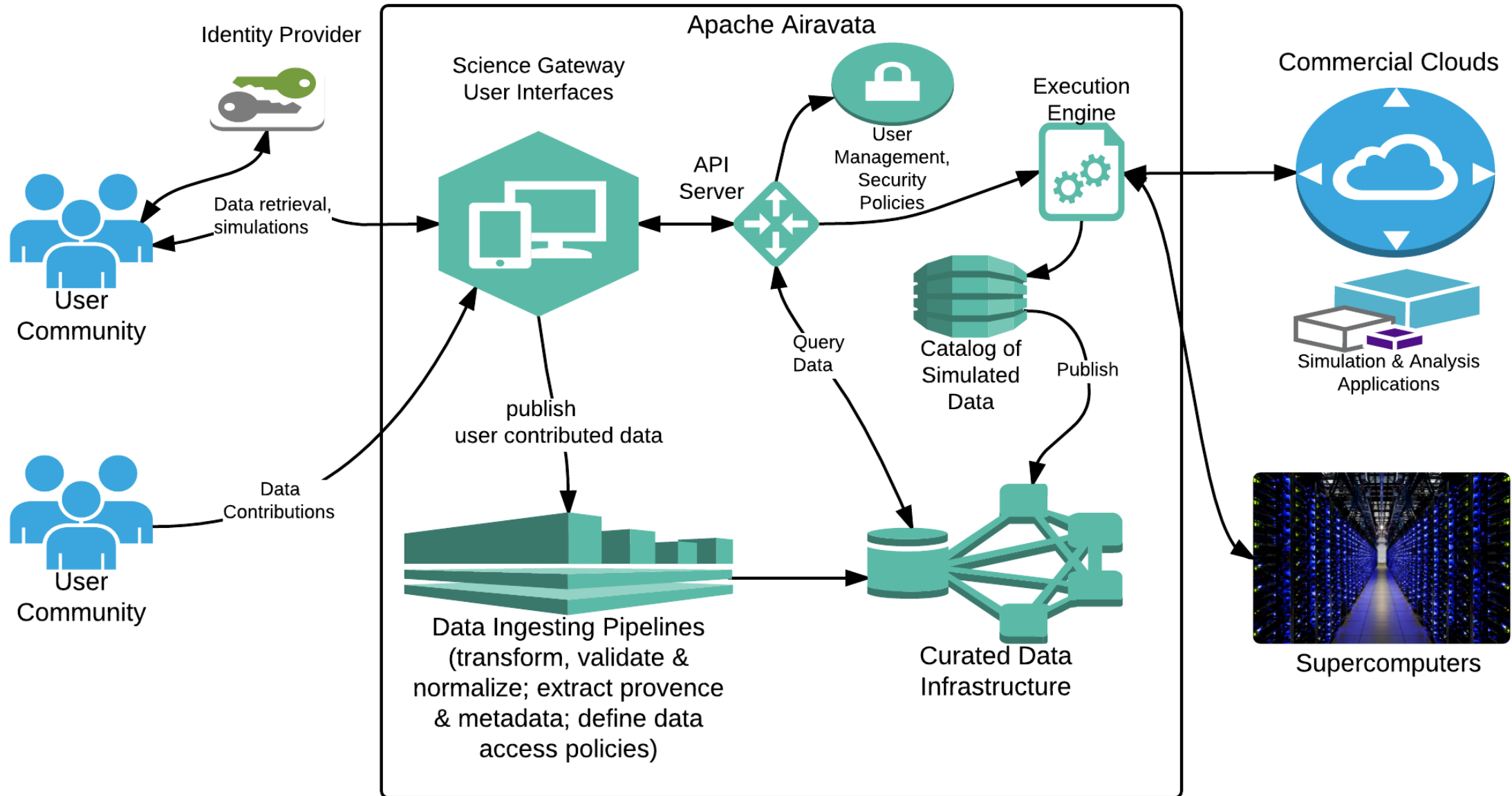
- “.....The stretch goal for LEAD was to begin ushering in this paradigm change. After 6 years, 415 papers and presentations, a mention of LEAD by Bill Gates at Supercomputing 2007, 9 PhD's awarded, 14 Master's degrees, and 1 Bachelor's degree;”

Generalizing LEAD and embracing
“Community over code”

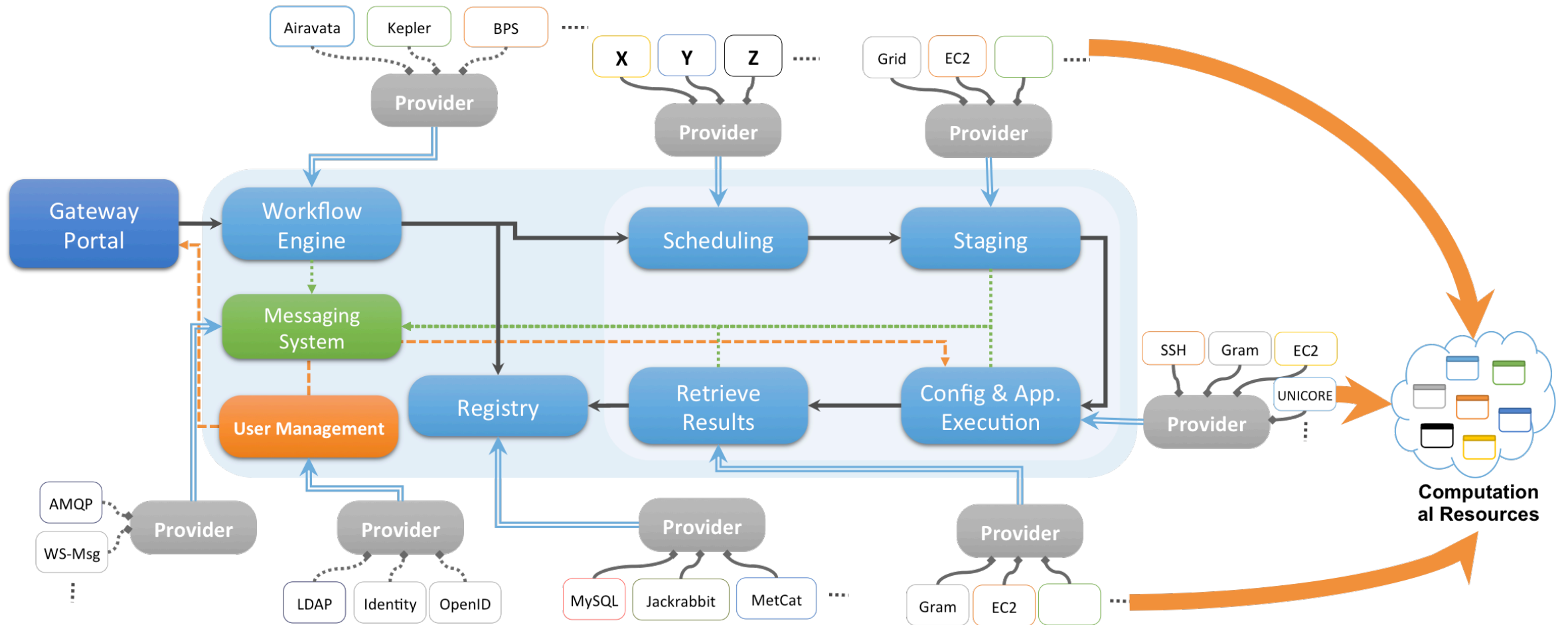
Towards a Generic SGW Hosted Platform



Apache Airavata Architecture

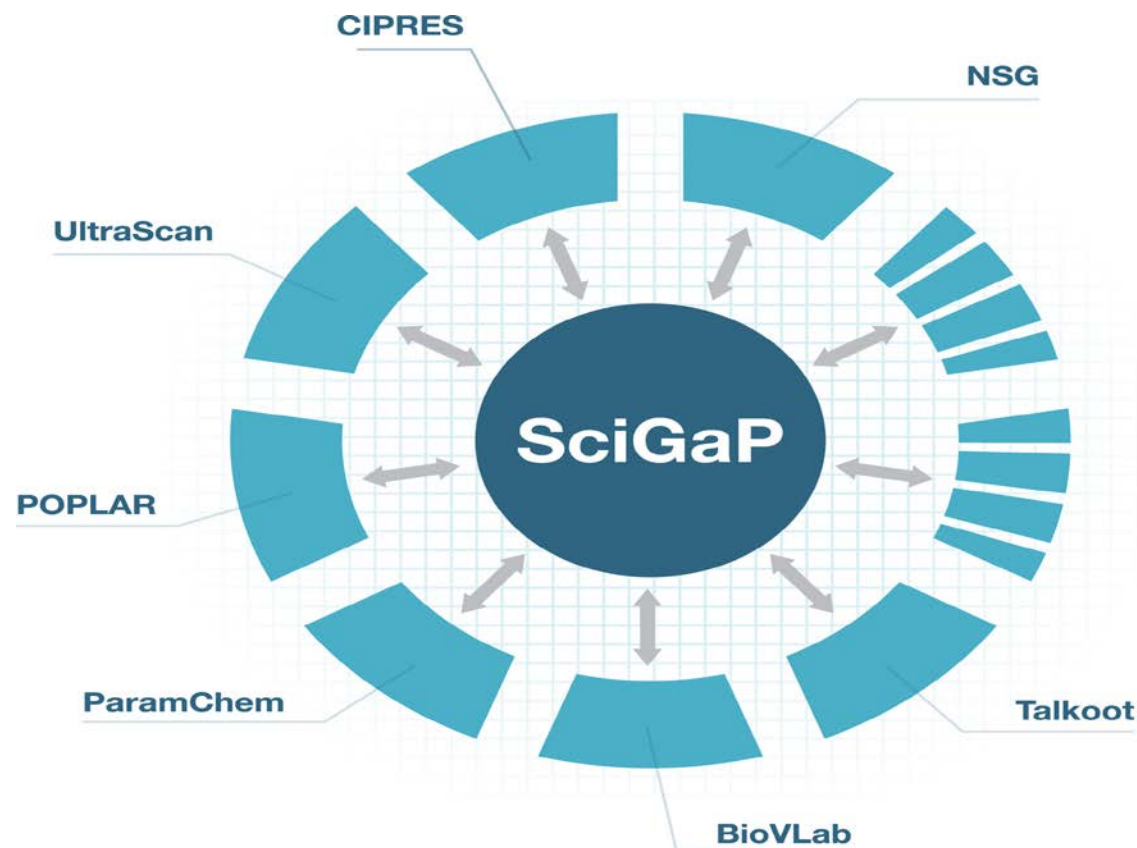


A Build vs Buy vs Collaborate Story



SciGaP Key Mission

Scale number of gateways without having to scale FTE's needed to support them.

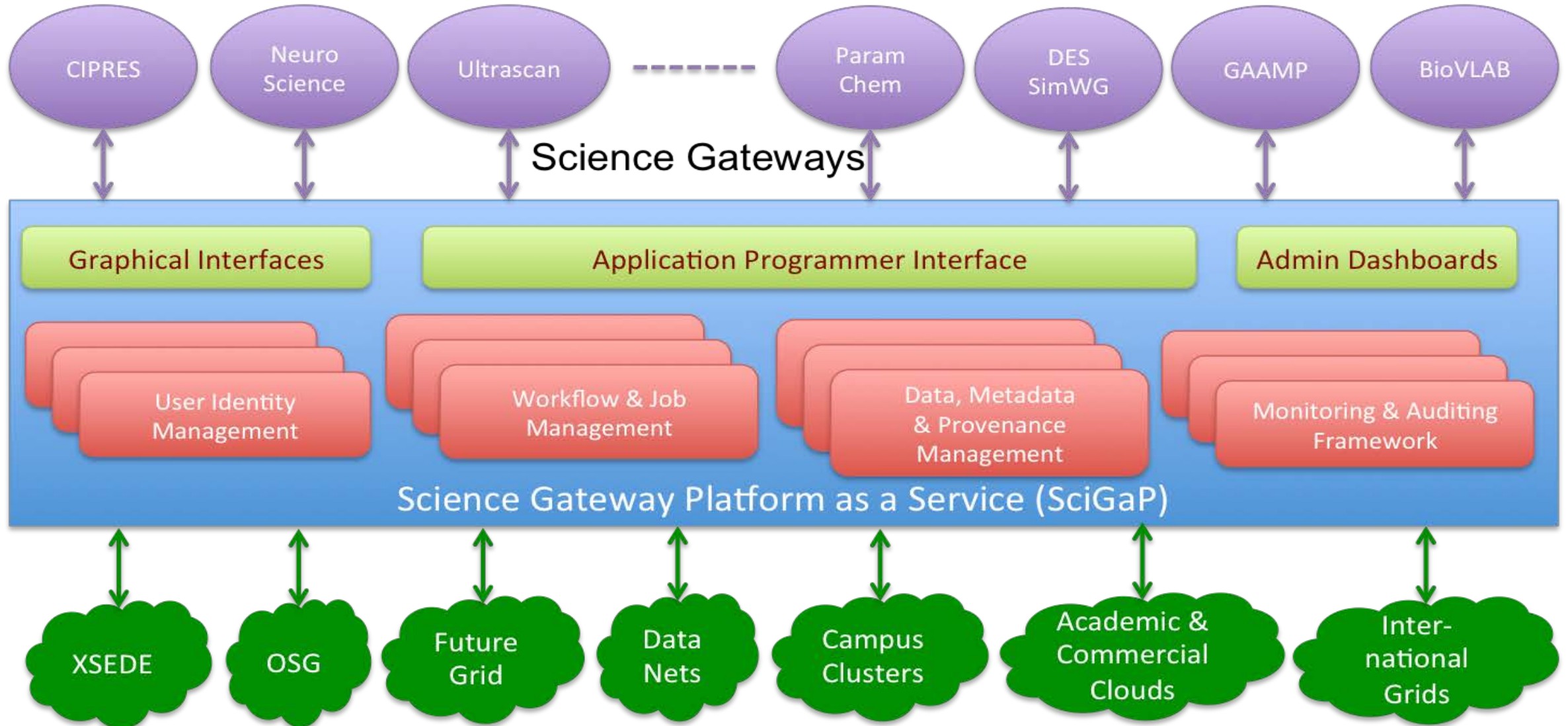


POWERED BY

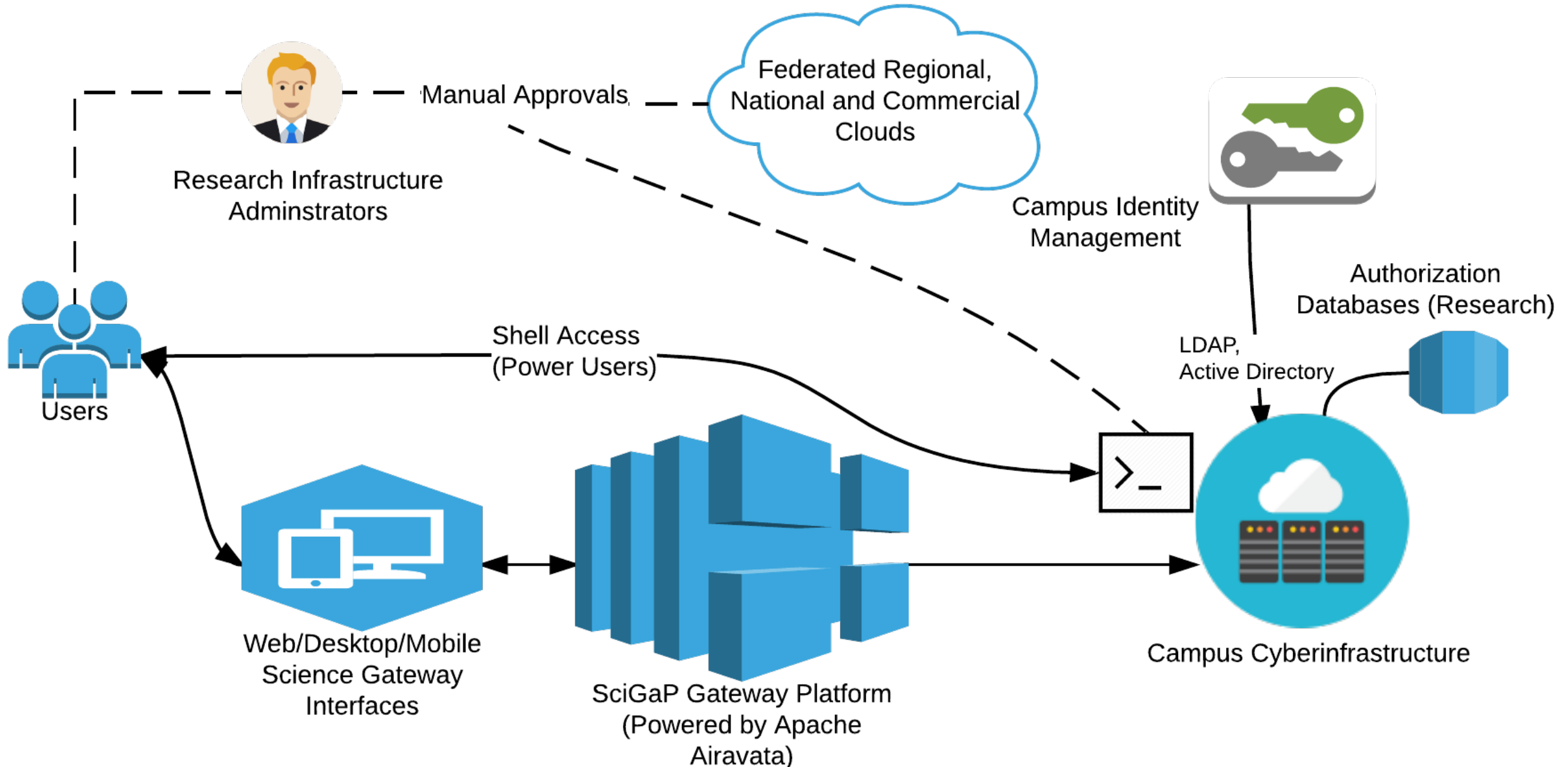


APACHE
AIRAVATA

Improve sustainability by converging on a single set of hosted infrastructure services



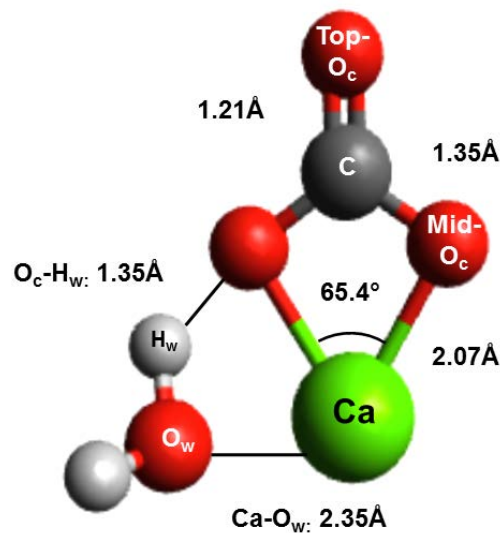
Single Campus Cyberinfrastructure



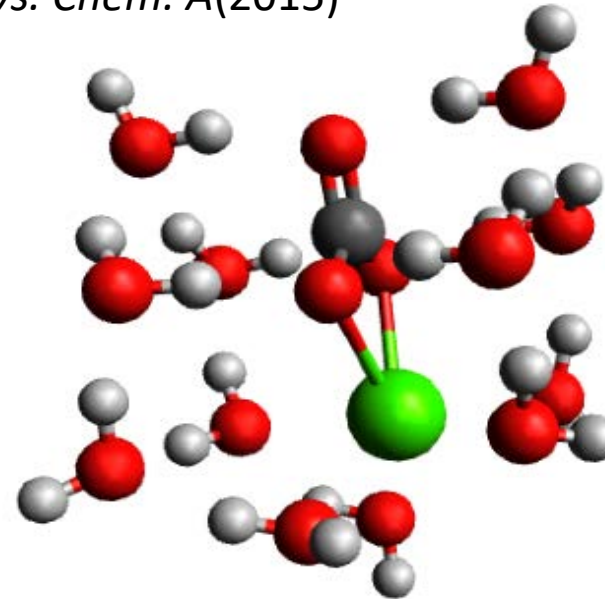
What is the chemistry of hydrated calcium carbonate?

- Bio-mineralization of skeletons and shells
- Geological CO₂ sequestration
- Cleanup of contaminated environments

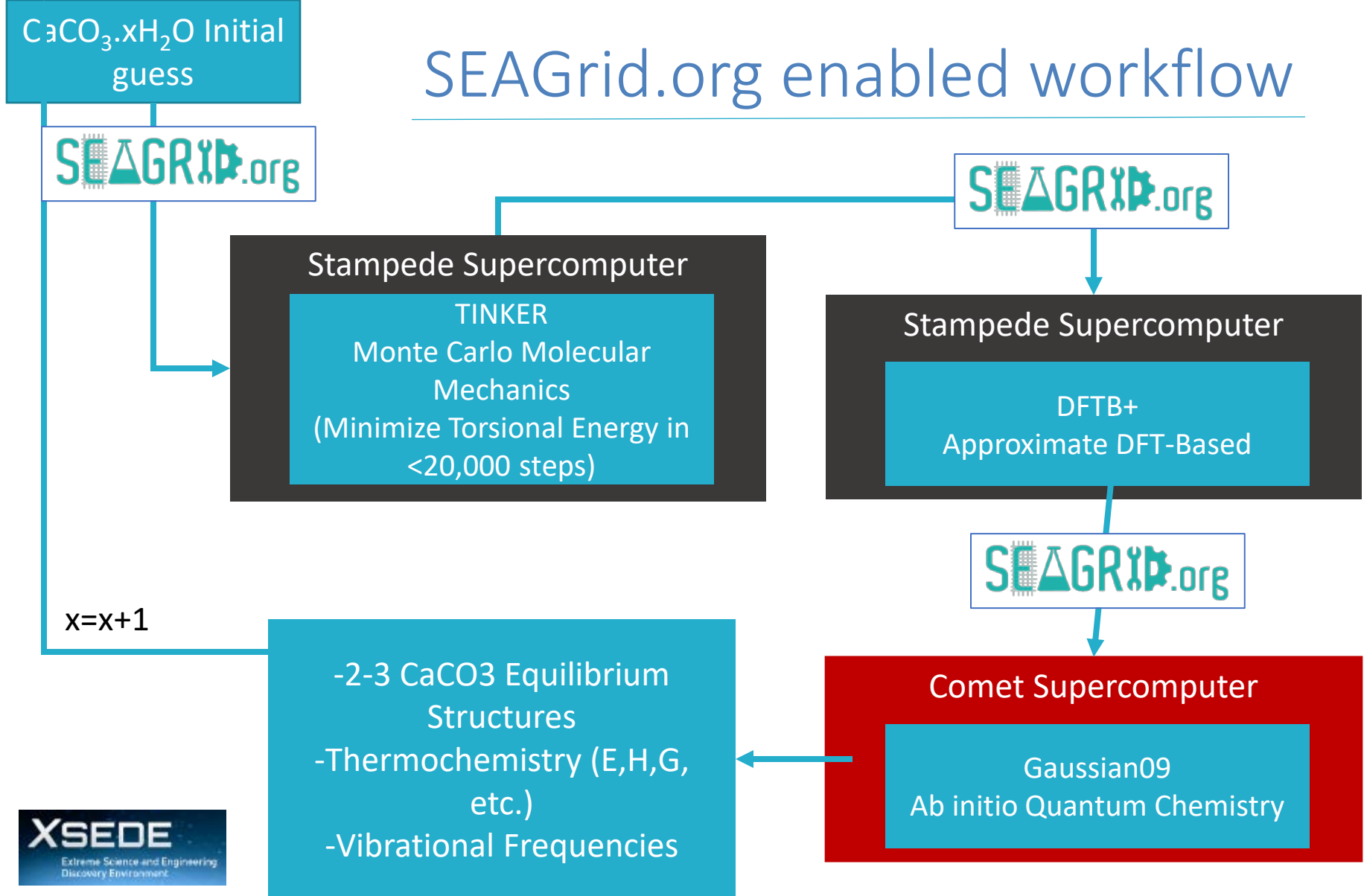
Lopez-Berganza, et al. *J Phys. Chem. A*(2015)



CaCO₃.1H₂O



CaCO₃.12H₂O



Lopez-Berganza, et al. *J Phys. Chem. A*(2015)

In the beginning, we had no services

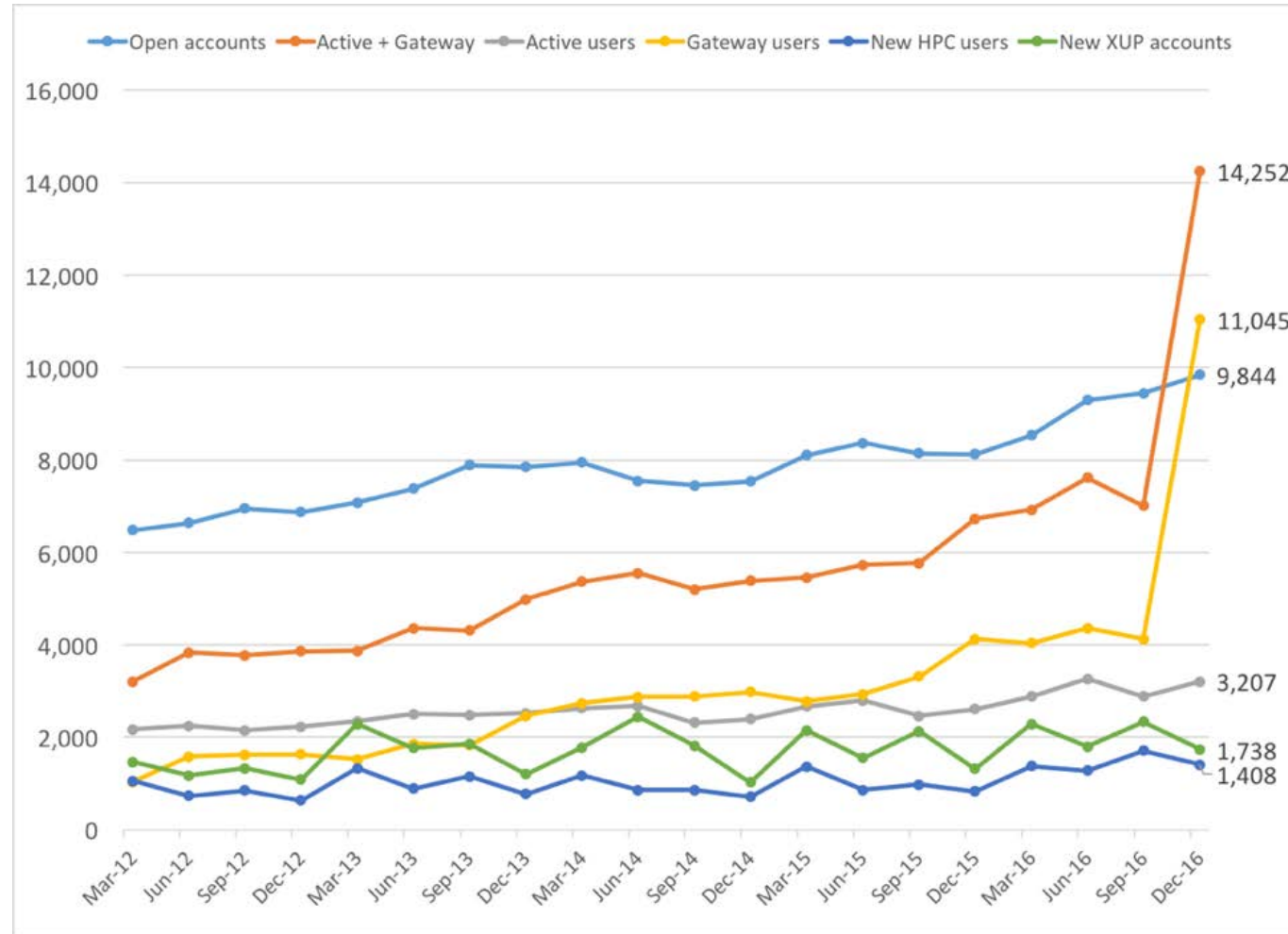
We paid science teams to help us develop them

<u>Science Gateway Prototype</u>	<u>Discipline</u>	<u>Science Partner(s)</u>	<u>TeraGrid Liaison</u>
Linked Environments for Atmospheric Discovery (LEAD)	Atmospheric	Droegemeier (OU)	Gannon (IU), Pennington (NCSA)
National Virtual Observatory (NVO)	Astronomy	Szalay (Johns Hopkins)	Williams (Caltech)
Network for Computational Nanotechnology (NCN) and "nanoHUB"	Nanotechnology	Lundstrum (PU)	Goasguen (PU)
Open Life Sciences Gateway	Biomedicine and Biology	Schneewind (UC), Osterman (Burnham/UCSD), DeLong (MIT), Dusko (INRA)	Stevens (UC/Argonne)
Biology and Biomedical Science Gateway	Biomedicine and Biology	Cunningham (Duke), Magnuson (UNC)	Reed (UNC), Blatecky (UNC)
Neutron Science Instrument Gateway	Physics	Cobb (ORNL)	Cobb (ORNL)
Grid Analysis Environment	High-Energy Physics	Newman (Caltech)	Bunn (Caltech)
Transportation System Decision Support	Homeland Security	Stephen Eubanks (LANL)	Beckman (Argonne)
Groundwater/Flood Modeling	Environmental	Wells (UT-Austin), Engel (ORNL)	Boisseau (TACC)
Science Grid [GrPhyN/ivDGL/Grid3]	Multiple	Pordes (FNAL), Huth (Harvard), Avery (Uflorida)	Foster (UC/Argonne), Kesselman (USC-ISI), Livny (UW)

Eventually we had a program

- And customers
- 2013, gateway users surpass command line users in XSEDE
- 2016, gateways now 77% of active XSEDE users

XSEDE users



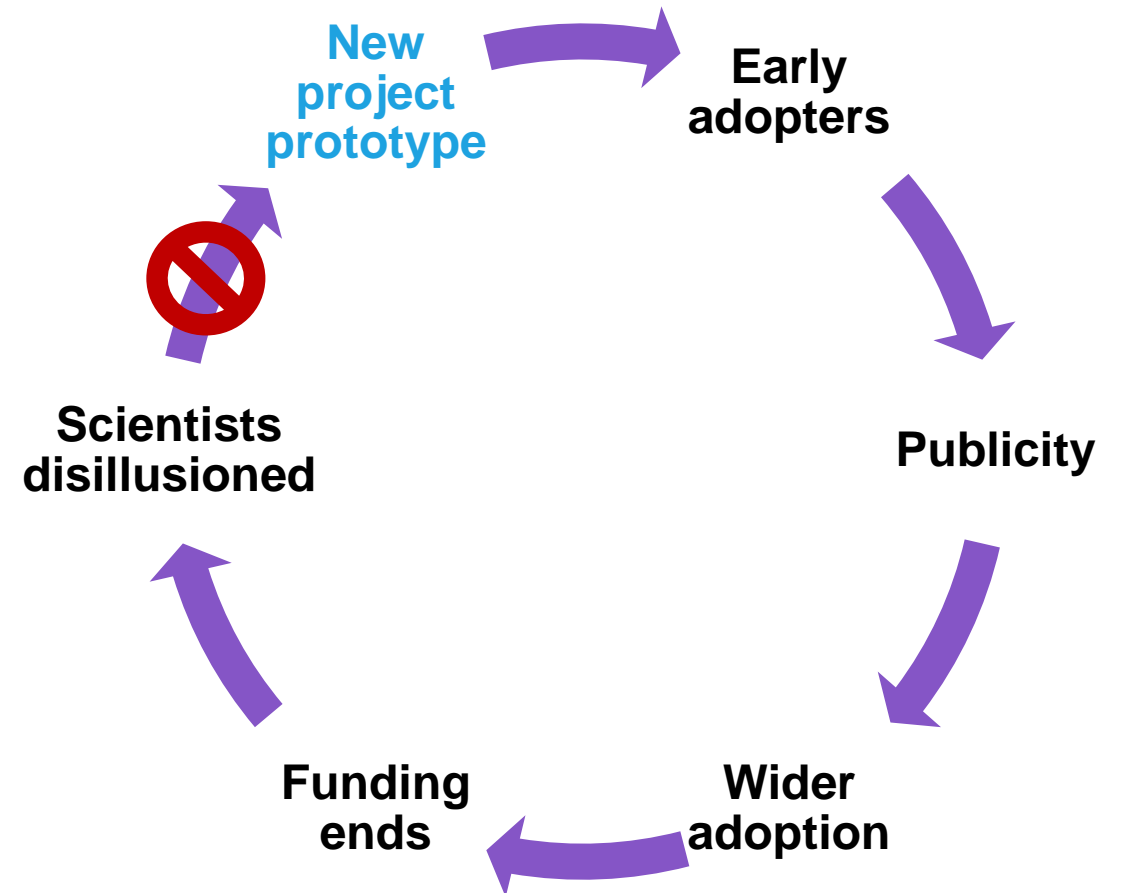
All users

Gateways

Despite many successes, we observed challenges

Gateways often funded as 3-year research projects

- Developers typically
 - work in isolation
 - must bridge to variety of resources
 - need building blocks in order to focus on higher-level functionality
 - struggle to secure sustainable funding



We studied the problem

And studied it some more

2009-2012
EAGER

Focus groups

2012-2015
Concept. phase

- More focus groups
- Survey with 5000 responses

2016
Software
Institute!

10+ year road to the birth of an institute

*Despite the technological progress of grid technology and deployment, **only a minority of the scientific, engineering, and education community use today's national computing infrastructure.** Our WIDE strategy addresses this situation by working directly with specific community leaders who are building **discipline-specific cyberinfrastructure capabilities** and resources for their communities.*

TeraGrid proposal, 2003



- “discipline-specific CI capabilities” = science gateways
 - **First example of community groups using supercomputers without individual identification**

5-year S2I2 Implementation phase award begins Aug, 2016

Press Release 16-088

NSF commits \$35 million to improve scientific software

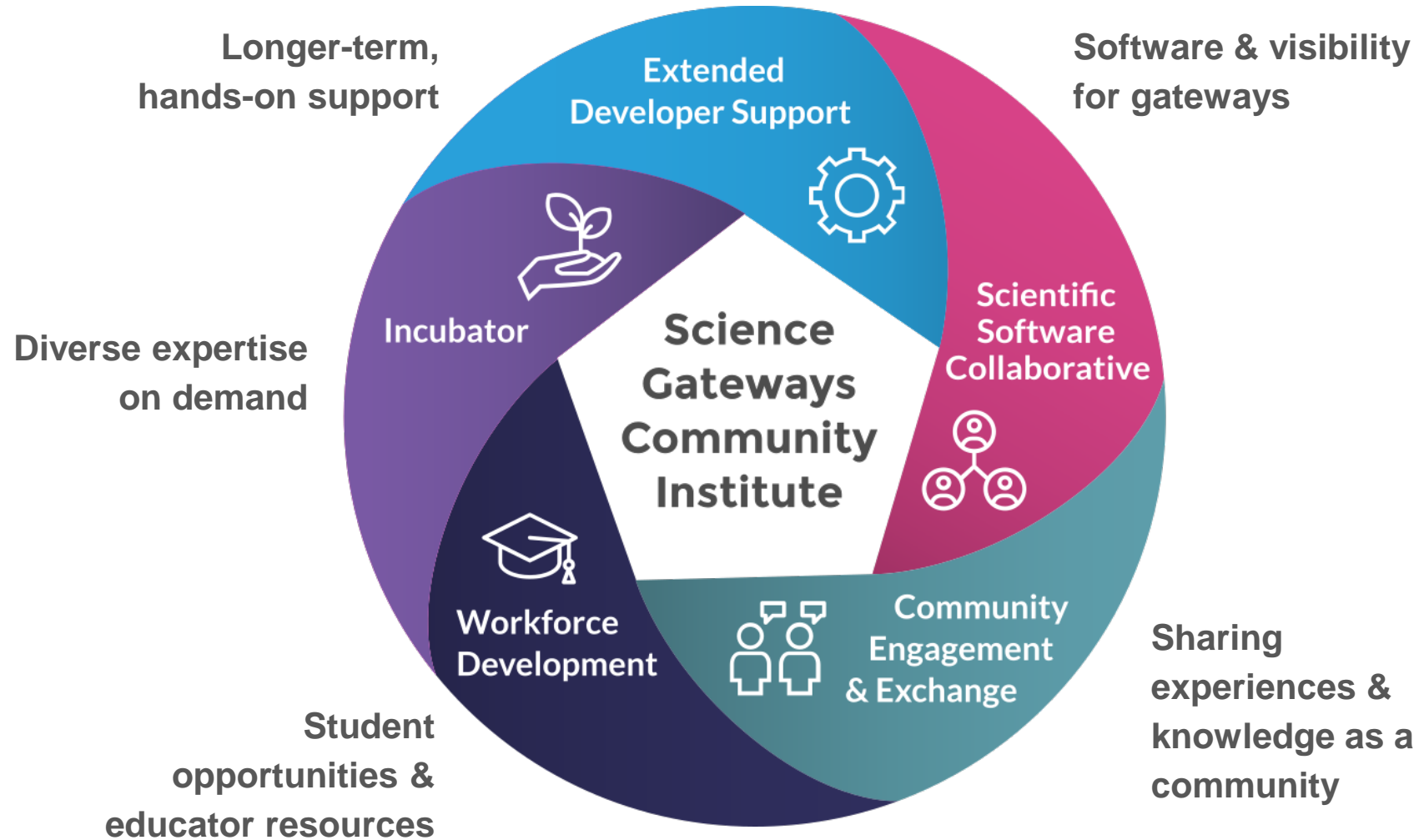
Science Gateways Community Institute

The second award, led by the University of California, San Diego, establishes the **Science Gateways** Community Institute, a multi-institutional consortium that will increase the capabilities, number and sustainability of **science gateways**. Gateways are mobile or web-based applications that provide broad access to the nation's shared cyberinfrastructure to scientists and citizens alike.

"Gateways foster collaborations and the exchange of ideas among researchers and can democratize access, providing broad access to resources sometimes unavailable to those who are not at leading research institutions," said Nancy Wilkins-Diehr, associate director of the San Diego Supercomputer Center and principal investigator for the project. "Sharing expertise about basic infrastructure allows developers to concentrate on the novel, the challenging, and the cutting-edge development needed by their specific user community."

Science Gateways Community Institute

Designed to help the community build gateways more effectively



Closing Thoughts

- Gateways translate resource-centric views into science-centric views.
- LEAD succeeded in ushering a “paradigm shift” from a technical perspective, but fell short in preparing the community.
 - LEAD is not currently functional as a virtual organization, but many of its constituent elements are. Any interest in resurrecting in its entirety?
- Gateways are developed by the community for the community. In addition to the longer term vision setting of this workshop, I would be thrilled to work with many of you to build “Gateways to Clouds” for research and education.
 - You can get “free” (payed by NSF) allocation of our time to help you in these efforts. We will help you build and operate gateways, but you are on your own to support the community.

More Information

- Science Gateways Research Center
 - <https://sgrc.iu.edu/>
- Apache Airavata Open Source Science Gateway Software
 - <http://airavata.apache.org/>
- Contact
 - Center email: sgrc-iu-group@iu.edu
 - Marlon Pierce: marpierc@iu.edu
 - Suresh Marru: smarru@iu.edu