Computation Institute

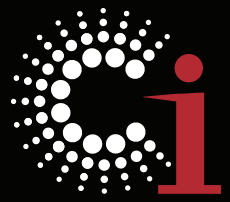
# Hybrid Strategies for Research Data Management

Vas Vasiliadis, Computation Institute

vas@ci.uchicago.edu

Argonne NATIONAL LABORATORY

computationinstitute.org



**The Computation Institute**

= UChicago + Argonne

= Cross-disciplinary nexus

= Home of the Research Cloud

Genetics

**High energy physics**

Molecular biology

**Climate change**

Cosmology

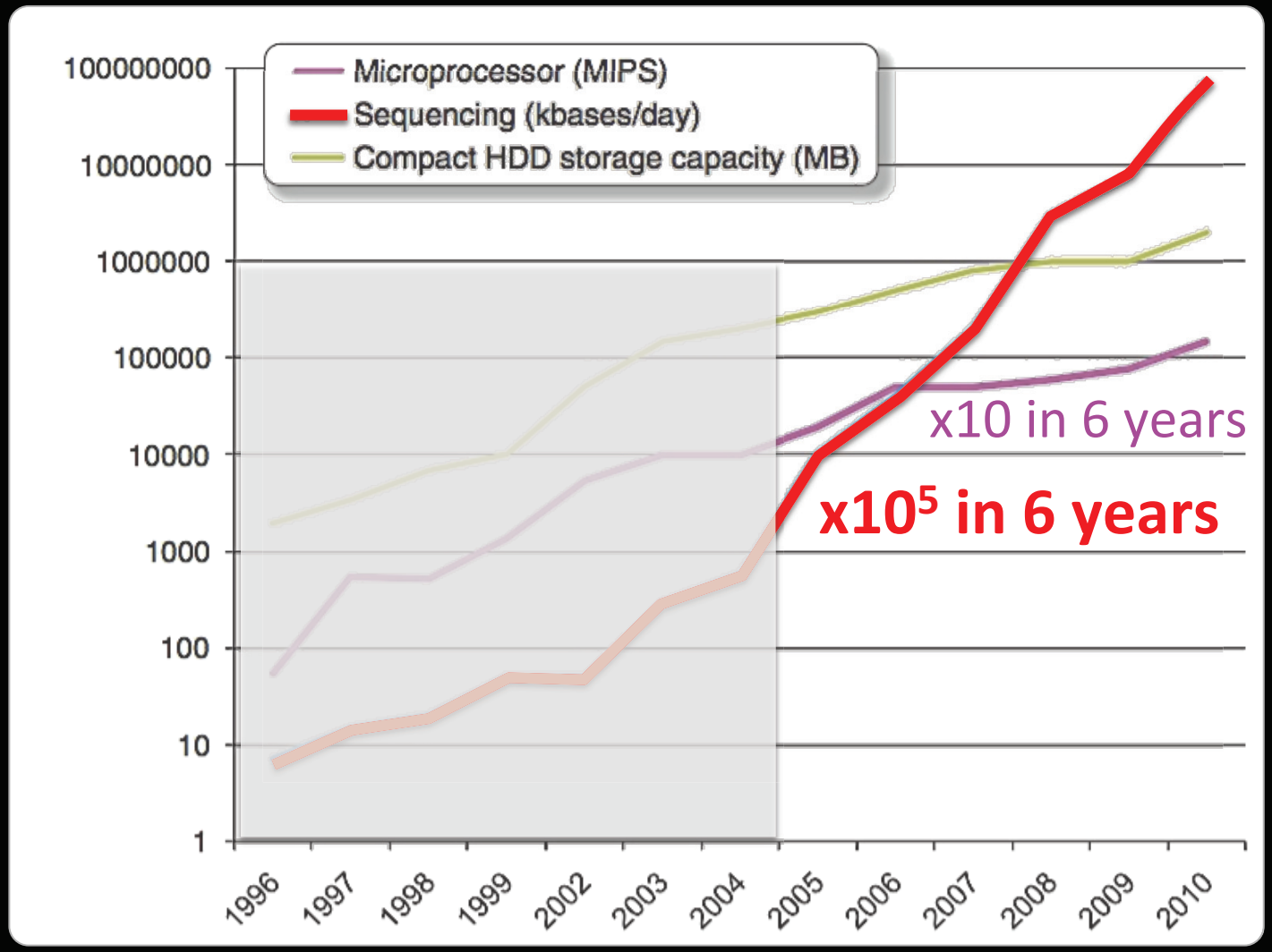
*Linguistics*

**Metagenomics**

Visual arts

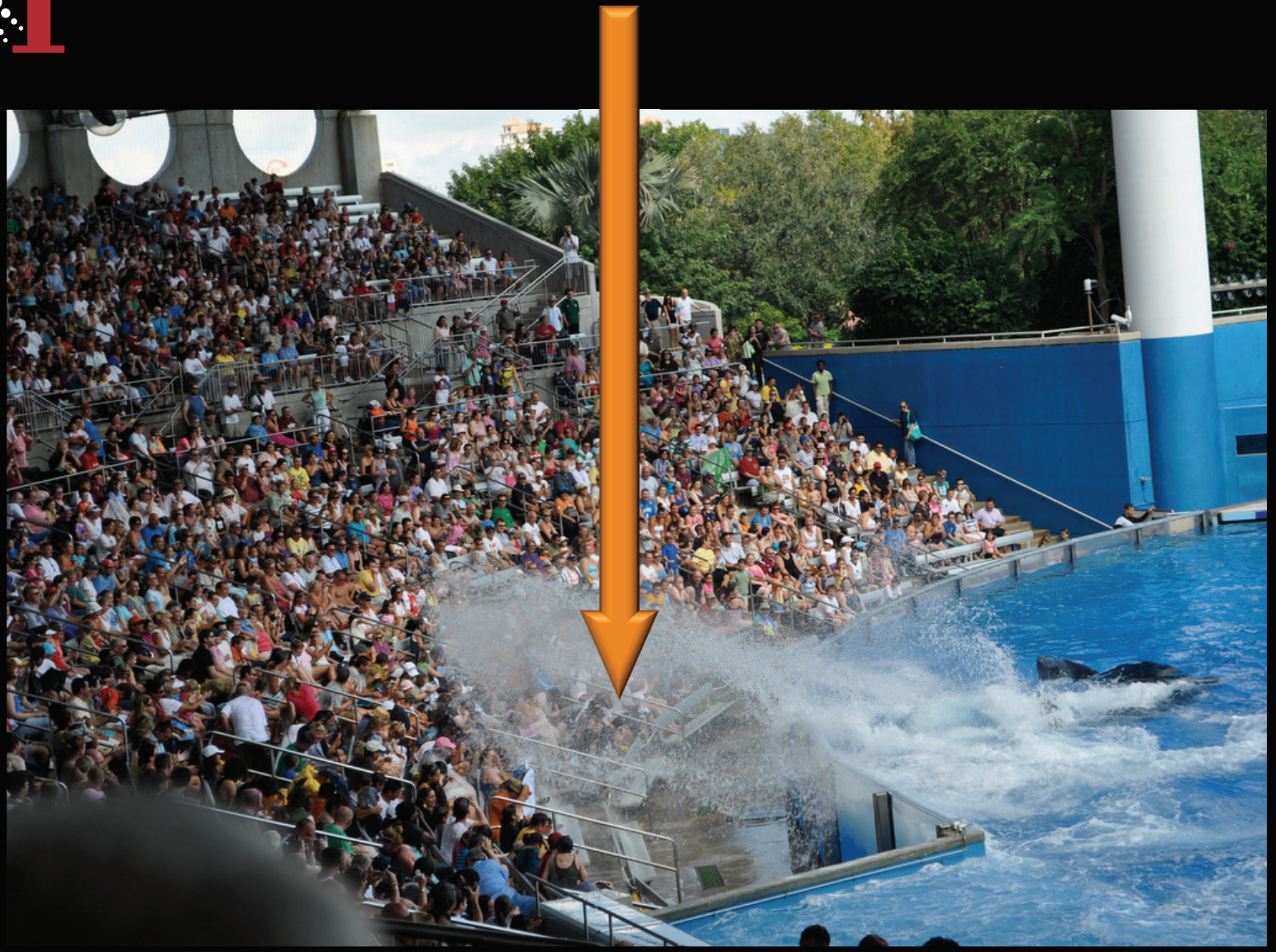
**Economics**

computationinstitute.org

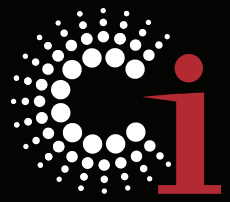

Microprocessor (MIPS)
Sequencing (kbases/day)
Compact HDD storage capacity (MB)
100000000
10000000
1000000
100000
10000
1000
100
10
1
1996
1997
1998
1999
2002
2003
2004
2005
2006
2007
2008
2009
2010
x10 in 6 years
x10^5 in 6 years


1 PB data in last experiment
Accessed by 800 scientists worldwide


# 1.2 PB of climate data
# Delivered to 23,000 users

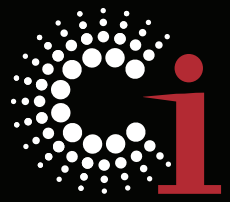


computationinstitute.org

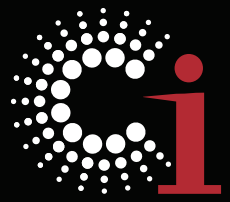


We have exceptional infrastructure for the 1%

How can the 99% manage this?



# What would a “dropbox for science” look like?



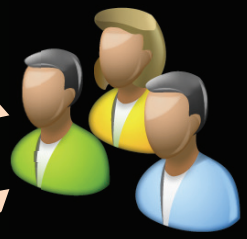
- Collect
- Move
- Replicate
- Share
- Analyze
- Catalog
- Publish
- Search
- Archive
- Backup

…among distributed research groups

Staging Store

Ingest Store

Registry



Analysis Store

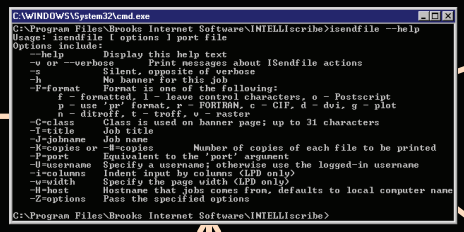
Community Store

Archive

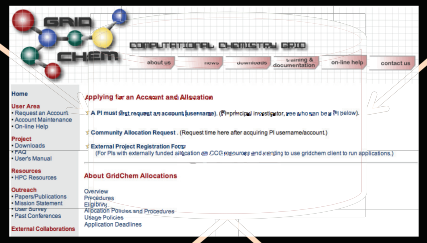
Mirror

computationinstitute.org

Staging Store




Registry


Archive

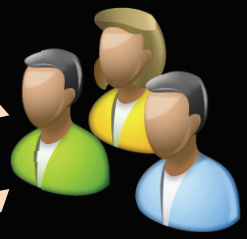
Mirror

computationinstitute.org

Staging Store

Ingest Store

Registry



Community Store

Analysis Store

Archive

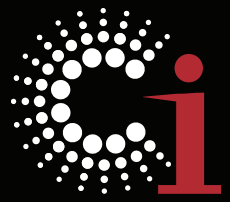
Mirror

computationinstitute.org

- Collect
- Move
- Replicate
- Share
- Analyze

- Catalog
- Publish
- Search
- Archive
- Backup

-as-a-Service



Security

Privacy

Reliability

Scalability

Control

computationinstitute.org

# A great user experience

# www.globusonline.org

## Research Data Management-as-a-Service

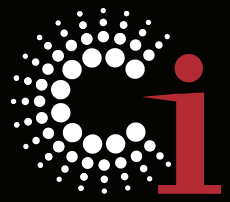
globus online

| | |
|---|---|
| **Globus Transfer** | SaaS |
| **Globus Storage** | |
| **Globus Collaborate** | |
| **Globus Catalog** | |
| **Globus Integrate** (Globus Nexus, Globus Connect) | PaaS |

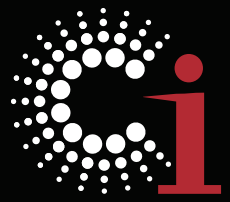
Supported by U.S. DEPARTMENT OF ENERGY · NSF · NATIONAL INSTITUTES OF HEALTH · THE UNIVERSITY OF CHICAGO · Argonne NATIONAL LABORATORY

# Communities using Globus


computationinstitute.org



# What does it mean for us as IT resource managers?



# installers → brokers

ITServices

amazon web services™

red cloud

Drupal

SDSC Cloud

JIRA

zendesk

ATLASSIAN CONFLUENCE®

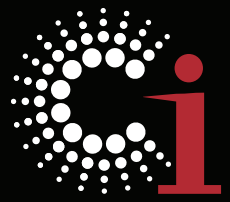
# developers → integrators


InCommon®


Shibboleth®


MyProxy
Credential Management Service

GSI-OpenSSH



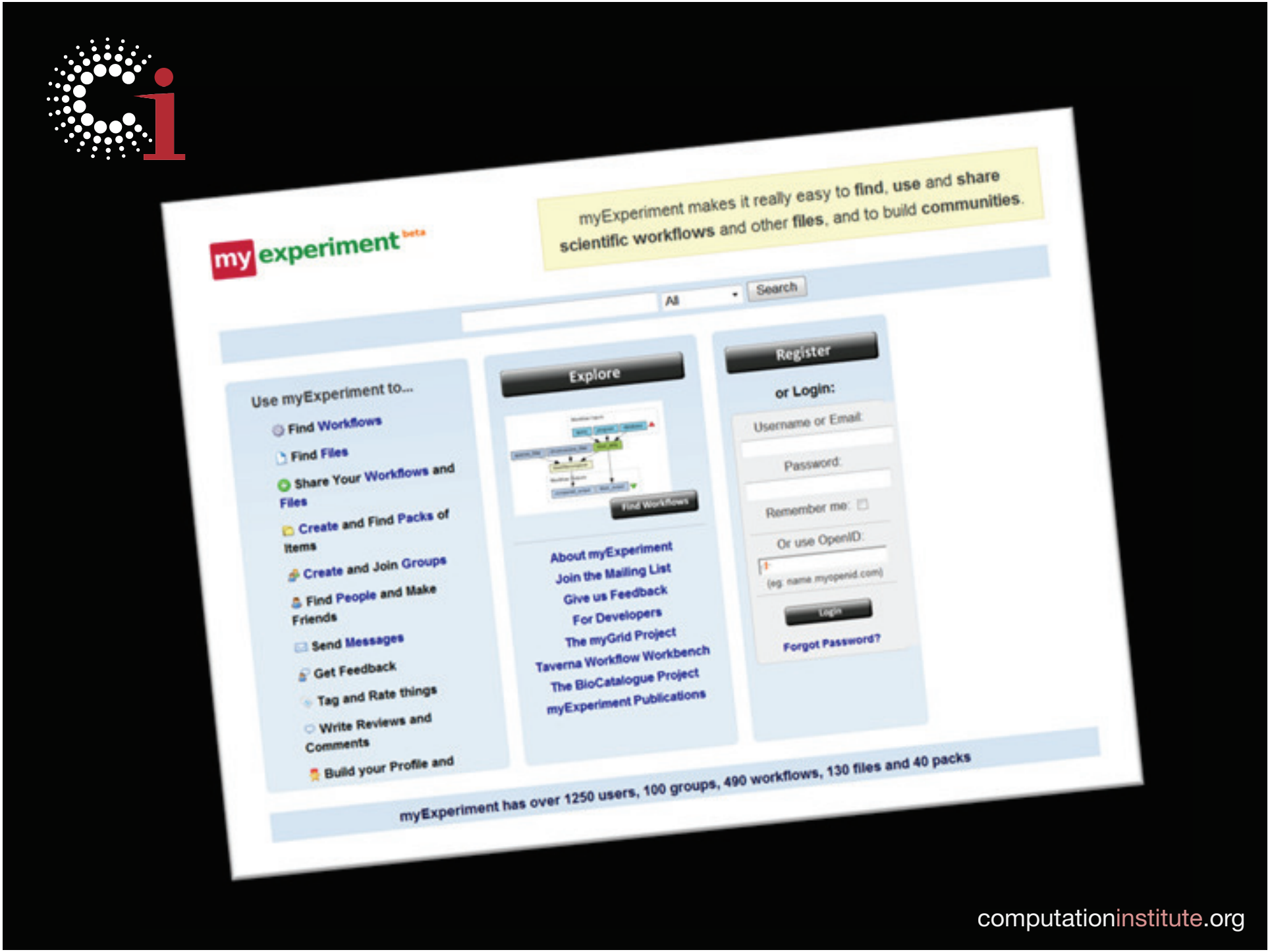
administrators → curators

(of the user experience)

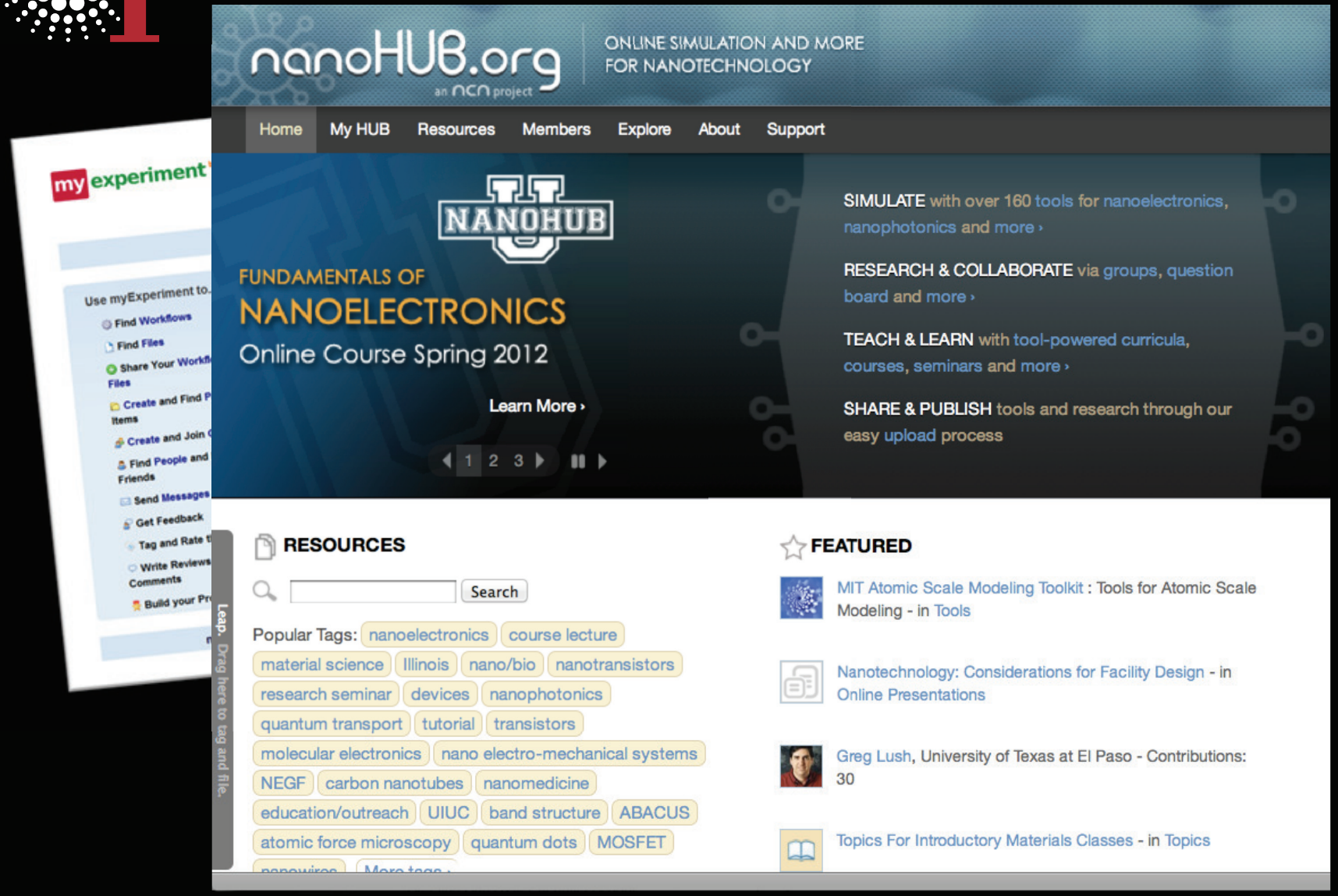
Cloud? What cloud?

**1 : 1 : 0**

UX : Dev : Ops

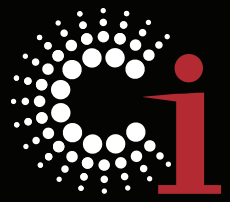



computationinstitute.org



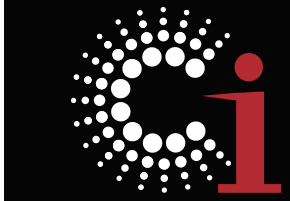

computationinstitute.org


computationinstitute.org



# Our vision for a 21st century cyberinfrastructure

To provide **more** capability for **more** people at **substantially lower cost** by creatively **aggregating** (“cloud”) and **federating** (“grid”) resources in a **hybrid world**



U.S. DEPARTMENT OF
ENERGY



THE UNIVERSITY OF
CHICAGO

Argonne
NATIONAL LABORATORY

CardioVascular Research Grid

computationinstitute.org