

# Cyberinfrastructure for Research: From campus growth to national trends (Part II of III)

**Craig Stewart**

**ORCID ID 0000-0003-2423-9019**

**Jetstream Principal Investigator**

**Executive Director, Indiana University Pervasive Technology Institute**

**Presented at Michigan State University, East Lansing, MI, 22 October 2015**

# XSEDE

**Extreme Science and Engineering  
Discovery Environment**



# So....

- Suppose you are on a good path to dealing with your campus CI issues.
- Needs always exceed supply in IT – for hardware, software, people, etc.
- What then?
- National level
  - XSEDE (eXtreme Science and Engineering Discovery Environment)
  - Open Science Grid
  - Incite (DOE)

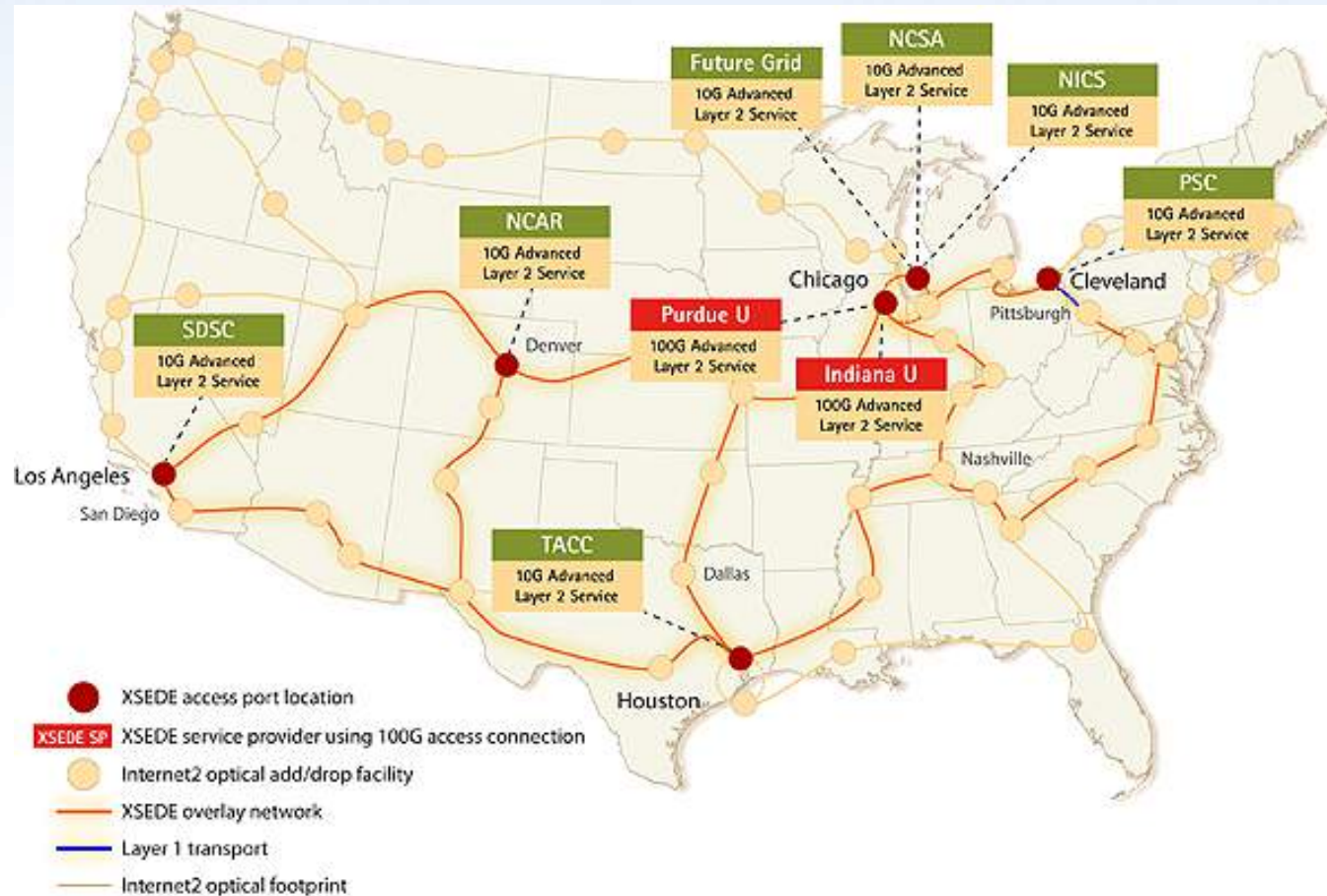


# XSEDE (xsede.org) is a national source of cyberinfrastructure resources

- Allocated
  - Cycles
  - Data storage
  - Support
  - Get help the first time you apply - [help@xsede.org](mailto:help@xsede.org) and/or your local campus champion
- Available to all (without allocations)
  - Globus Transfer
  - Training & curriculum materials
  - National Cyberinfrastructure Integration



# XSEDE – a national cyberinfrastructure instrument



From [xsede.org](http://xsede.org)



# XSEDE

# XSEDE resources

- Managing your own systems
  - XCBC (XSEDE Compatible Basic Cluster)
- Systems for you to use
  - \*Jetstream coming in 2016
  - Bridges coming in 2016
  - Comet available now
  - \*Wrangler available now
- Consulting Help
  - XSEDE ECSS
- NCGAS (National Center for Genome Analysis Support)
- All funded by federal government and available via allocations



# Campus Bridging – XSEDE National Integration Toolkit (XNIT)

- Software tools to:
  - Make it easier for your local systems administrators to manage your local clusters.
  - Make it easier for you to make your local clusters more consistent with systems supported by XSEDE (diversity of names and partners notwithstanding, there is a lot of consistency across systems).
  - Subscribe to the tools you want. Ignore the ones you don't.
  - Or build a cluster from scratch.



The logo for Jetstream features the word "Jetstream" in a bold, italicized, red sans-serif font. A light blue swoosh underline starts under the 'J' and extends across the word. Below the main text, the tagline "A national science & engineering cloud" is written in a smaller, white, sans-serif font.

# **Jetstream**

**A national science & engineering cloud**

**funded by the National Science Foundation  
Award #ACI-1445604**

Dashboard

Images

Favorites

My Images

Projects

Cloud Providers

Quotas

Settings

# Search Images

Search by App Images, Tag, OS, and more

Popular Searches: [R](#) [Bisque](#) [NGS](#) [Community: Astrophysics](#)


Quick Sort:  Popularity  Recency  Rating

[Advanced Search Options](#)

Quick Filter:

View as:

## Popular Images from All Communities




**Math Kernel Library**

[blas](#) [fft](#) [fortran](#) [lapack](#)

Community: Mathematics

52 likes, 0 dislikes, 7 comments




**RNASeq Analysis Tools**

[bowtie2](#) [blast](#) [blat](#) [edgeR](#)

[R](#) [maseq](#) [tophat2](#)

Community: Biology

30 likes, 2 dislikes, 4 comments




**Atmospheric Dispersion Modeling**

[aermod](#) [aermet](#) [aermap](#)

Community: Atmospheric Sciences

20 likes, 0 dislikes, 0 comments






**MrBayes with TreeMix**

[bayesian inference](#) [mrbayes](#)

[treemix](#)

Community: Phylogenetics

25 likes, 1 dislike, 10 comments





# What is Jetstream?

- NSF's first cloud for science and engineering research across all areas of activity supported by the NSF.
- Jetstream will be a user-friendly cloud environment designed to give researchers and research students access to interactive computing and data analysis resources “on demand.”
- It will provide a user-selectable library of virtual machines from which users can select to do their research.
- Software creators and researchers will also be able to create their own customized virtual machines or their own “private computing system” within Jetstream.
- It will enable many discoveries across disciplines such as biology, atmospheric science, economics, network science, observational astronomy, and social sciences.
- Two especially important biology platforms will be supported - iPlant and Galaxy.



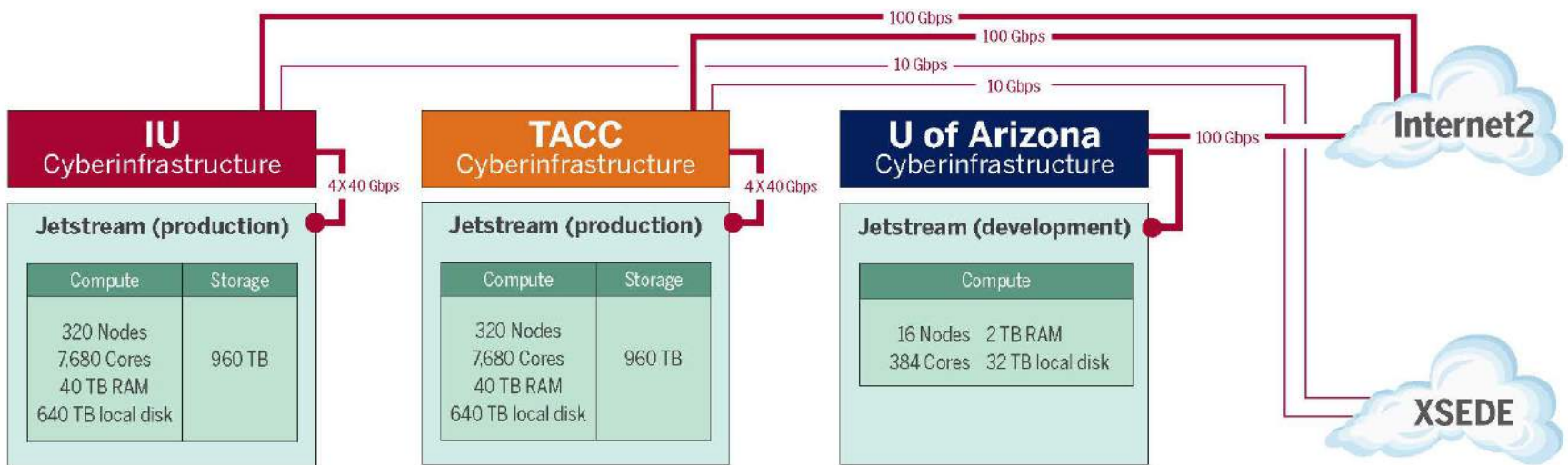
XSEDE

# What does the name mean? And is it really a cloud?

- Name
  - In the atmosphere the Jetstream lies at the border of two different air masses.
  - The Jetstream system stands at the border of the existing NSF-funded XD program and advanced cyberinfrastructure resources and users who have not previously used such NSF-funded infrastructure.
- Yep, it's really a cloud, or at least a cloud environment (one could quibble over the definition of cloud vis-à-vis expansibility). Software layers:
  - Atmosphere interface
  - KVM
  - OpenStack
  - CentOS Linux



# Jetstream System Diagram



# What is Wrangler?

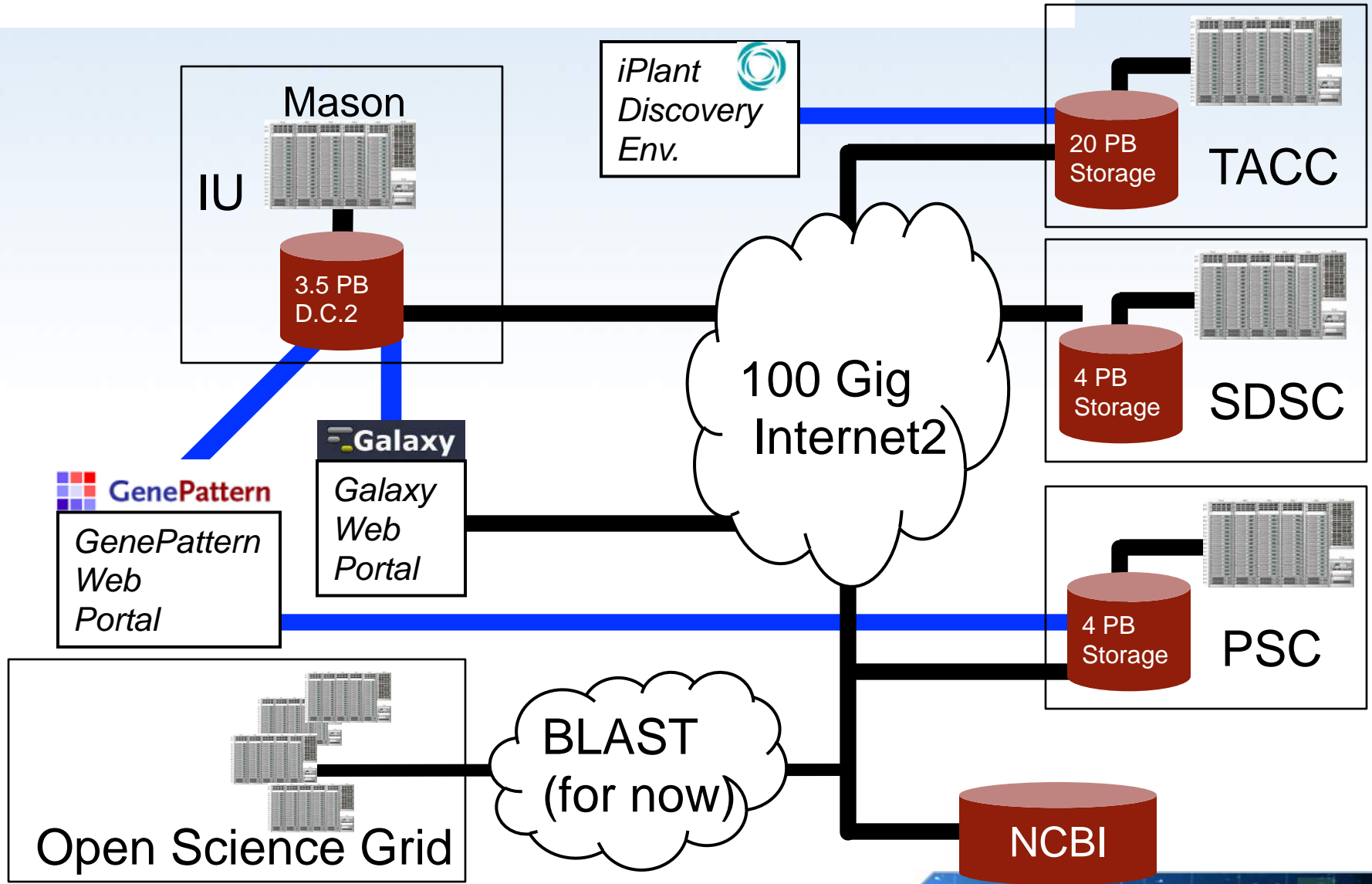
- A new data-intensive supercomputing system
- Built from the ground up for “data intensive” applications
- Massive, replicated, secure, high performance data storage (10PB each at Indiana and TACC)
- A large-scale flash storage tier for analytics, with bandwidth of 1TB/s and 250M IOPS (6x faster than Stampede)
- Embedded processing of more than 3,000 processors cores for data analysis
- Flexible support for a wide range of data workflows, including those using Hadoop and databases. Integration with Globus Online services for rapid and reliable data transfer and sharing.



XSEDE

# National Center for Genome Analysis Support

## XD Resources



# XSEDE

# XSEDE ECSS- Extended Collaborative Support Services

- Expertise available in many fields
  - over 50 expertise areas
- Sometimes serve as an intellectual commons bringing disparate research groups together for increased productivity
  - e.g. among users running large-scale genomics calculations
- Have often significantly increased user productivity and user capability
  - e.g. median code speedup 2.25x, highest speedup 126x, over 200 live training/outreach events in PY3.

# But you do have to apply for resources

- *Resources are available for use in research projects by faculty, staff, and students. Resources are also available in support of classroom education. You do not need current NSF funding to apply.*
- Go to [xsede.org](https://www.xsede.org) and make a portal account (easy).
- For resources allocated through XSEDE, fill out application form at <https://www.xsede.org/allocations>
- Current round of allocation requests deadline 22 October!
- Help from
  - [help@xsede.org](mailto:help@xsede.org)
  - [campusbridging@xsede.org](mailto:campusbridging@xsede.org)
  - [jethelp@iu.edu](mailto:jethelp@iu.edu)
  - [ncgas@iu.edu](mailto:ncgas@iu.edu)
- Ask for help asking for help!



# Jetstream Collaborators

- University of Chicago - Globus
- Arizona University – iPlant
- Johns Hopkins University and Penn State University
- Cornell University –Ms. Susan Mehringer, Lead. Cornell® Virtual Workshops about Jetstream and applications running on Jetstream.
- University of Arkansas at Pine Bluff – Dr. Jesse Walker, lead. cybersecurity education, Minority Serving Education outreach
- University of Hawaii – Dr. Gwen Jacobs, lead. EPSCoR early adopter/user. Jacobs will chair Science Advisory Board.
- National Snow and Ice Data Center (NSIDC) – Dr. Ron Weaver, lead. Data retrieval from NSIDC, application integration with ice-sheet analysis applications
- University of North Carolina, Odum Center –Dr. Thomas Carsey, lead. Data retrieval from Dataverse Network
- National Center for Genome Analysis at Indiana University – providing genome analysis software. Includes TACC, PSC, and SDSC as partners.



XSEDE





UC San Diego

**SDSC**  
SAN DIEGO SUPERCOMPUTER CENTER



INDIANA UNIVERSITY



*This work supported by the National Science Foundation, award ACI-1341698.*



**XSEDE**

# NCGAS Partners



National Institutes  
of Health



**TEXAS ADVANCED COMPUTING CENTER**

*Powering Discoveries That Change The World*



**XSEDE**

Extreme Science and Engineering  
Discovery Environment



Open Science Grid

**SDSC**  
SAN DIEGO SUPERCOMPUTER CENTER



**XSEDE**

# Acknowledgments & Disclaimers

- Thanks to Nick Nystrom of the Pittsburgh Supercomputing Center for information about the new Bridges System. Bridges is supported by NSF award 1445606
- Thanks to Richard Moore of the San Diego Supercomputer Center for information about Comet. Comet is supported by NSF award 1341698
- Thanks to Daniel Stanzione of the Texas Advanced Computing Center for slides about Wrangler. Wrangler is supported by NSF award 1341711
- Jetstream is supported by NSF award 1445604 (Craig Stewart, PI)
- XSEDE is supported by NSF award 1053575 (John Towns, UIUC, PI)
- This work was also supported by the Indiana University Pervasive Technology Institute, which was initiated with major funding from the Lilly Endowment, Inc.
- Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF) or other supporting organizations.



# License Terms

Craig A. Stewart. 2015. Cyberinfrastructure for Research: From campus growth to national trends (Part II of III). Presentation. Presented at Michigan State University, East Lansing, MI. 22 October 2015. <http://hdl.handle.net/2022/20443>

Items indicated with a © are under copyright and used here with permission. Such items may not be reused without permission from the holder of copyright except where license terms noted on a slide permit reuse.

Except where otherwise noted, contents of this presentation are copyright 2014 by the Trustees of Indiana University.

This document is released under the Creative Commons Attribution 3.0 Unported license (<http://creativecommons.org/licenses/by/3.0/>). This license includes the following terms: You are free to share – to copy, distribute and transmit the work and to remix – to adapt the work under the following conditions: attribution – you must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work). For any reuse or distribution, you must make clear to others the license terms of this work.

This research was supported in part by the National Science Foundation through Award ACI-1445604. This research was supported in part by the Indiana University Pervasive Technology Institute, which was established with the assistance of a major award from the Lilly Endowment, Inc. Opinions presented here are those of the author(s) and do not necessarily represent the views of the NSF, IUPTI, IU, or the Lilly Endowment, Inc.

