Open Science Forum, April 26, 2017

IVMOOC on Jetstream & Open XD Metrics on Demand Value Analytics

CNS, IUNI, UITS

INDIANA UNIVERSITY BLOOMINGTON
Open Science Forum

Wed, Apr 26th 2017 at 4:00 PM
SSRC Grand Hall (Woodburn Hall 200)

Jetstream, funded by NSF and led by the Indiana University Pervasive Technology Institute (PTI), adds cloud-based computation to the national cyberinfrastructure (https://jetstream-cloud.org). IUNI collaborated with the Cyberinfrastructure for Network Science Center to deploy the Network Workbench tool on Jetstream for use by IVMOOC students (http://ivmooc.cns.iu.edu).

The "Open XD Metrics on Demand Value Analytics" NSF project, directed by Matthew Link (UITS), aims to measure and visualize the impact of campus-financed cyberinfrastructure (CI) on progress in science. Resulting visual analytics modules will be added to the existing CI metrics tool eXtreme Data Metrics on Demand (XDMoD) to present a view of financial, collaboration, and publication data, showing “return-on-investment” metrics in relation to CI usage.
New Ventures in Research, Engineering, and Educational Computing.

George Turner, Chief Systems Architect
Research Technologies, UITS, Indiana University

Open Science Forum, SSRC Grand Hall
Indiana University Bloomington, IN
26 April 2017
What is Jetstream?

- **User-friendly**, widely accessible cloud environment

  - **User-selectable library** of preconfigured virtual machines; no need for system administration skills.

  - **Programmatic API access** to implement modern cloud computing techniques

http://jetstream-cloud.org/
What is Jetstream?

- **Reproducibility**: store, publish via IU Scholarworks (DOI)
- **Cloudy**: clouds are more the just virtual machines (VM)
  - Old way: robust (expensive) infrastructure, weak (cheap) software
  - Cloudy way: commodity infrastructure, robust software
  - Cows, not pets: pets take great amount of care, feeding, and you name them; cows you intend to have high turnover and you give them numbers.
- **Primary goal** is to expand the user base of NSF’s eXtreme Digital (XD) program resources beyond the current community of users.
What is Jetstream? (cont)

“Long tail” of the Science

Large HPC systems requiring sophisticated programming skills

few

Capable users

many

http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604
What is Jetstream? (cont)

- **Software layers**
  - **Atmosphere** web interface
    - library of images, generic, domain specific
    - simplify VM administration
  - **Openstack**: software tools for building and managing cloud computing platforms for public and private clouds.
  - **KVM** hypervisor: what the VMs run on
  - **Ceph**: storage platform that stores data on a single distributed computer cluster, and provides interfaces for object-, block- and file-level storage.
  - **Operating systems**: CentOS, Ubuntu, Windows?
  - **Applications**: e.g. software developed by the domain specialist, gateways, etc.
Jetstream System Overview

Jetstream (production)
- Compute: 320 Nodes, 7,680 Cores, 40 TB RAM, 640 TB local disk
- Storage: 960 TB

Jetstream (production)
- Compute: 320 Nodes, 7,680 Cores, 40 TB RAM, 640 TB local disk
- Storage: 960 TB

Jetstream (development)
- Compute: 16 Nodes, 2 TB RAM, 384 Cores, 32 TB local disk
- Storage: 960 TB

funded by the National Science Foundation
Award #ACI-1445604

http://jetstream-cloud.org/
# Production Cloud Hardware (per site)

<table>
<thead>
<tr>
<th>Hardware</th>
<th>Number</th>
<th>Specifications</th>
<th>Function (IU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dell PowerEdge M630 blades</td>
<td>320</td>
<td>2X Intel E5-2680v3 “Haswell” 24 cores @ 2.5 GHz 128 GB RAM 2 TB local disk</td>
<td>Compute hosts OpenStack services</td>
</tr>
<tr>
<td>Dell PowerEdge R630 1U server</td>
<td>7</td>
<td>2X Intel E5-2680v3 “Haswell” 24 cores @ 2.5 GHz 128 GB RAM 2 TB local disk</td>
<td>Cluster management High Availability Databases RabbitMQ</td>
</tr>
<tr>
<td>Dell PowerEdge R730xd 2U servers</td>
<td>20</td>
<td>2X Intel E5-2680v3 “Haswell” 24 cores @ 2.5 GHz 64 GB RAM 48 TB storage for Ceph pool</td>
<td>~1 PB Ceph storage</td>
</tr>
<tr>
<td>Dell S6000-ON network switches</td>
<td>9</td>
<td>32+2 40 Gb/s ports</td>
<td>Top of Rack Spine</td>
</tr>
</tbody>
</table>

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604
## VM Instance Sizes (Flavors)

<table>
<thead>
<tr>
<th>Instance Type</th>
<th>vCPUs</th>
<th>RAM(GB)</th>
<th>Storage(GB)</th>
<th>Instances/Node</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiny</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>46</td>
</tr>
<tr>
<td>Small</td>
<td>2</td>
<td>4</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>Medium</td>
<td>6</td>
<td>16</td>
<td>60</td>
<td>7</td>
</tr>
<tr>
<td>Large</td>
<td>10</td>
<td>30</td>
<td>120/60*</td>
<td>4</td>
</tr>
<tr>
<td>X-Large</td>
<td>22</td>
<td>60</td>
<td>240/60*</td>
<td>2</td>
</tr>
<tr>
<td>XX-Large</td>
<td>44</td>
<td>120</td>
<td>480/60*</td>
<td>1</td>
</tr>
</tbody>
</table>

Node config: dual Intel E-2680v3 “Haswell”, 24 physical cores/node @ 2.5 GHz, 128 GB RAM, dual 1 TB local disks.

* Effective 29-Mar-2017
What is an Instance

Diagram showing a single box divided into sections labeled RAM, CPU, and Storage.
Instance & Volumes

Attach a persistent volume to an ephemeral instance

RAM

CPU

Storage

Volume

funded by the National Science Foundation
Award #ACI-1445604

http://jetstream-cloud.org/
Instance & Volumes

Dettach the persistent volume to an ephemeral instance
The instance is gone but the volume persists.
Terms

- Image: a file on a storage device
- Instance: a running
- Volumes: persistent storage device
- Flavors: the size of instance, #cpu, amt of RAM, amt of storage
- Host: hardware
- Guest:
Jetstream’s Atmosphere Interface
(no login required at this point)

https://use.jetstream-cloud.org/
Jetstream’s Atmosphere Interface

(Pick identity provider)
Jetstream’s Atmosphere Interface

(Authenticate)
Jetstream’s Atmosphere Interface

(user’s home space)
OpenStack Organization

funded by the National Science Foundation
Award #ACI-1445604
OpenStack Projects

http://www.openstack.org/software/project-navigator/

funded by the National Science Foundation
Award #ACI-1445604
Openstack Projects  ...the core services

<table>
<thead>
<tr>
<th>Service</th>
<th>Name</th>
<th>Adoption</th>
<th>Maturity</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identity</td>
<td>Keystone</td>
<td>96%</td>
<td>7/8</td>
<td>5 yrs</td>
</tr>
<tr>
<td>Images</td>
<td>Glance</td>
<td>95%</td>
<td>6/8</td>
<td>7 yrs</td>
</tr>
<tr>
<td>Block device</td>
<td>Cinder</td>
<td>88%</td>
<td>7/8</td>
<td>5 yrs</td>
</tr>
<tr>
<td>Networking</td>
<td>Neutron</td>
<td>93%</td>
<td>7/8</td>
<td>5 yrs</td>
</tr>
<tr>
<td>Compute</td>
<td>Nova</td>
<td>95%</td>
<td>8/8</td>
<td>7 yrs</td>
</tr>
<tr>
<td>Object device</td>
<td>Swift</td>
<td>52%</td>
<td>7/8</td>
<td>7 yrs</td>
</tr>
</tbody>
</table>

https://www.openstack.org/software/project-navigator/
Openstack Projects ...some other services

<table>
<thead>
<tr>
<th>Service</th>
<th>Name</th>
<th>Adoption</th>
<th>Maturity</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dashboard</td>
<td>Horizon</td>
<td>87%</td>
<td>6/8</td>
<td>5 yrs</td>
</tr>
<tr>
<td>Telemetry</td>
<td>Ceilometer</td>
<td>55%</td>
<td>1/8</td>
<td>4 yrs</td>
</tr>
<tr>
<td>Database</td>
<td>Trove</td>
<td>13%</td>
<td>3/8</td>
<td>3 yrs</td>
</tr>
<tr>
<td>Orchestration</td>
<td>Heat</td>
<td>67%</td>
<td>6/8</td>
<td>4 yrs</td>
</tr>
<tr>
<td>Provisioning</td>
<td>Ironic</td>
<td>17%</td>
<td>2/8</td>
<td>3 yrs</td>
</tr>
</tbody>
</table>

https://www.openstack.org/software/project-navigator/
# Openstack Projects ...some other services

<table>
<thead>
<tr>
<th>Service</th>
<th>Name</th>
<th>Adoption</th>
<th>Maturity</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Map/Reduce</td>
<td>Sahara</td>
<td>10%</td>
<td>3/8</td>
<td>3 yrs</td>
</tr>
<tr>
<td>Shared Filesystems</td>
<td>Manila</td>
<td>14%</td>
<td>5/8</td>
<td>3 yrs</td>
</tr>
<tr>
<td>DNS Service</td>
<td>Designate</td>
<td>16%</td>
<td>3/8</td>
<td>3 yrs</td>
</tr>
<tr>
<td>Containers</td>
<td>Magnum</td>
<td>11%</td>
<td>2/8</td>
<td>2 yrs</td>
</tr>
<tr>
<td>Application Catalog</td>
<td>Murano</td>
<td>11%</td>
<td>1/8</td>
<td>2 yrs</td>
</tr>
</tbody>
</table>

https://www.openstack.org/software/project-navigator/
Just for fun: Happy Cluster – Mad Cluster
Infrared image of Jetstream
Internet

HA Proxy

Controller 1

Controller 2

Controller 3

funded by the National Science Foundation
Award #ACI-1445604

http://jetstream-cloud.org/
High Availability layout for the databases

Spine 1
Top of Rack
Database 1
Database 2
Load Balancer 2

Spine 2
Top of Rack
Database 3
Load Balancer 1
Management
Glance - Cinder - Ceph
OpenStack Overview

Client -> Keystone -> Glance

Keystone -> Ceph

Keystone -> Nova

Nova -> Ceph

Nova -> Cinder

Cinder -> Compute

Glance -> Compute

Ceph -> Compute

Token
VXLAN Packet

- Ethernet
- IP/UDP
- Ethernet
- IP
Neutron Networking

Hypervisor
VM
Bridge
VXLAN

Network node
VXLAN
Bridge
NetNS
Bridge

Internet

funded by the National Science Foundation
Award #ACI-1445604
Network Topology (cont.)

Sixteen blades per chassis
Two switches per chassis

Chassis

To Top of Rack Switches

10 Gb/s per blade

10 GB/s per blade

2 x 40 Gb/s Inter-switch links

10 Gb/s

40 Gb/s

LAG
Network Topology (cont.)

Chassis to Top of Rack

Four chassis per rack

Two switches per chassis

40 Gb/s

LAG
Network Topology (cont.)

100 Gb/s uplink to Internet2

Two spines tie into two network blades in the datacenter switch

Seven racks tie into the two spine switches

40 Gb/s LAG

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

http://jetstream-cloud.org/
How do we onboard users onto Jetstream?

• An XSEDE User Portal (XUP) account is required. They are free! Get one at https://portal.xsede.org

• Read the Allocations Overview - https://portal.xsede.org/allocations-overview

• Write a successful allocation request – start with a Startup or Education request - https://portal.xsede.org/successful-requests
Jetstream Information Sources

- Jetstream: [https://use.jetstream-cloud.org/](https://use.jetstream-cloud.org/)

- XSEDE User Portal is required to actually login: [https://portal.xsede.org](https://portal.xsede.org)


- Configuration management: [https://github.com/jetstream-cloud/Jetstream-Salt-States](https://github.com/jetstream-cloud/Jetstream-Salt-States)
Questions?

Project website: http://jetstream-cloud.org/
Project email: jethelp@iu.edu
Direct email: jomlowe@iu.edu, turnerg@iu.edu

License Terms

• Jetstream is supported by NSF award 1445604 (Craig Stewart, IU, PI)
• XSEDE is supported by NSF award 1053575 (John Towns, UIUC, PI)
• 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, IUPUI, IU, or the Lilly Endowment, Inc.
• 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 2015 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.
Open Science Forum, April 26, 2017

IVMOOC on Jetstream

CNS, IUNI, UITS

INDIANA UNIVERSITY BLOOMINGTON
The Information Visualization MOOC
ivmooc.cns.iu.edu

Students from ~100 countries
370+ faculty members
#ivmooc
Part 1: Theory and Hands-On

**Session 1** – Workflow Design and Visualization Framework

**Session 2** – “When:” Temporal Data

**Session 3** – “Where:” Geospatial Data

**Session 4** – “What:” Topical Data

**Mid-Term**

**Session 5** – “With Whom:” Trees

**Session 6** – “With Whom:” Networks

**Session 7** – Dynamic Visualizations and Deployment

**Final Exam**

**Part 2: Students work in teams on client projects.**

Final grade is based on Class Participation (10%), Midterm (30%), Final Exam (30%), and Client Project (30%)
Books Used in the IVMOOC

Teaches timely knowledge:
Advanced algorithms, tools, and hands-on workflows.

Teaches timeless knowledge:
Visualization framework—exemplified using generic visualization examples and pioneering visualizations.
Sci2 Tool Interface Components
Download tool for free at http://sci2.cns.iu.edu
IVMOOC 2017: Using NSF XSEDE and Jetstream to Run Data Analysis and Visualization Workflows in the Cloud

Katy Borner, 2017.04.20

In Spring 2017, IVMOOC students are invited to beta-test a novel cloud computing setup that supports low- to high-bandwidth users in running more compute intensive data analysis and visualization workflows in the cloud. Specifically, students will create a user account for the Extreme Science and Engineering Discovery Environment (XSEDE) and then use the Jetstream cloud-computing environment to run specific workflows using the Network Workbench Tool.
Hi students,

We wanted to let you know that there is a new assignment posted today that students can complete to earn an extra 2% to their final grade. ([https://iu.instructure.com/courses/1595331/quizess/2271360](https://iu.instructure.com/courses/1595331/quizess/2271360))

To earn the extra credit, students will have to create an account with XSEDE cloud computing program, and then run a compute instance on the Jet Stream platform to test running a workflow in the cloud environment.

To earn the extra credit, you will need to create your XSEDE account and submit your username to us by Monday April 24 at noon, then we’ll link your account to Jetstream. Then from April 25-April 27th at noon, you will need to create an instance on the Jetstream, run a workflow, and submit your result to us.

Best,

Michael, Katy, and Andreas
Account Setup

In order to use the IVMOOC Virtual Desktop via the Jetstream cloud environment, you need to get an XSEDE user account and this account needs to be linked to the IVMOOC cloud instance so that you can use this unique resource.

- Create an XSEDE user account via [XSEDE User Portal](#).
- Submit your XSEDE user name via Canvas by 4/24, noon.
  (XSEDE user names were batch-added to IVMOOC VM)
- All IVMOOC students who submitted their XSEDE user name on time gained access to IVMOOC instance on Jetstream by 4/25.
Create to IVMOOC Instance on Jetstream

- After 4/25, noon, go to https://use.jetstream-cloud.org/application/images
- Login to use Jetstream using your existing organizational login, select “Indiana University” and use DUO.
- Select “Launch New Instance”
- Search for “IVMOOC” or directly go to https://use.jetstream-cloud.org/application/images/366
- Click on “Launch” in top right.
- Keep default values for the instance but increase “Instance Size” to m1.small (2 CPUs, 4096 GB memory, 20 GB disk). Click “Launch Instance” in lower right, see next slide.
- Wait until Status is “Active” then click on Instance with Name “IVMOOC.”
- Wait until “Activity” is N/A (about 1-2 mins). The VW is now ready for usage.
- Reload page. “Open Web Desktop” should now be visible in lower right. Click on it to open virtual desktop in web browser.
- Confirm “Use default config.”
Run NWB Tool

- On Web Desktop, double click Network Workbench (NWB) to run the tool.
- Run “Model > Hypergrid” and generate a network with 10,000 nodes and a Maximum degree of each node: 4.
- With the Hypergrid Network Model being selected in the Data Manager, run “Analysis > Network Analysis Toolkit (NAT).”
- Record the “Average degree” for nodes and submit value via Canvas.
- Feel free to try out other workflows.

Suspend/Stop IVMOOC Instance on Jetstream

When done, go back to Jetstream web interface and click on “Suspend” or “Stop” in Actions list on right:
• Logging allows for capturing all user actions.
• Supports student evaluation and guidance.
• Helps identify frequency of dataset/algorithu usage per branch of science in support of future tool development.
Open Science Forum, April 26, 2017

Open XD Metrics on Demand Value Analytics

CNS, IUNI, UITS

INDIANA UNIVERSITY BLOOMINGTON
Matt Link
Associate Vice President (Acting)
Director, Systems
Research Technologies, Pervasive Technology Institute
Office of the Vice President for IT, Indiana University

Supported by the National Science Foundation
Developed by the University at Buffalo Center for Computational Research

- Comprehensive resource management for HPC systems
- Provide detailed operational and usage data
- Support optimization of HPC resource utilization
- Facilitate planning and analysis
- Used for XSEDE metrics
XDMoD – Value Analytics

- NSF eager award to Indiana University and University at Buffalo
- Collaboration between the Center for Computational Research at the University at Buffalo, and the Pervasive Technology Institute and Center for Network Science at Indiana University
- Enables academic institutions to better understand Return On Investment (ROI) on advanced Cyberinfrastructure (CI)
- Shows the value of:
  - Fostering collaboration
  - Supporting scientific publications
  - Show relationship between campus CI and external grant funding
- Local XDMoD installation required – all data kept locally²
User Interface
VA module will be integrated into local version.
Grant data
"Since Big Red’s installation in 2006, users of that system were PIs or project directors on a total of $253 million in external funding, which includes $65.4M in facilities and administration funds. In addition to the research dollars flowing into the University, the ability to leverage IU’s leadership in research cyberinfrastructure has aided many other grants awarded to IU."

–2010
HPC and Biomedical Usage...

No use of Research CI
39%

Biomedical Software Usage
25%

HPC Facility Usage
18%

Screen shots of PTI developed statistics tool.

PTI: Funding analytics

RT Stats

Biochemistry/Molecular Biology (Indianapolis)
School Of Medicine

<table>
<thead>
<tr>
<th>Biochem Auxiliary Services</th>
<th>Bioinformatics</th>
<th>Biomedical Genomics Program</th>
<th>Diabetes Basic Science Res Ctr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdul Sater, Zahi A</td>
<td>Arthur, Jack W</td>
<td>Cerabona, Donna</td>
<td>Conteh, Abbas</td>
</tr>
<tr>
<td>Craven, Kelly</td>
<td>Edenberg, Howard J</td>
<td>Folck, Anthony F</td>
<td>Fox, Melanie J</td>
</tr>
<tr>
<td>Fusakio, Michael</td>
<td>Gendron, Jaimie M</td>
<td>Georgiadis, Millie M</td>
<td>Heyen, Joshua W</td>
</tr>
<tr>
<td>Heyerdahl, Darcy</td>
<td>Hoang, Quyen Q</td>
<td>Huang, Fei</td>
<td>Hunter, Gerald Q</td>
</tr>
</tbody>
</table>

Grants
24 totaling $6,056,352.00

<table>
<thead>
<tr>
<th>Grant</th>
<th>Project</th>
<th>Status</th>
<th>Dates</th>
<th>Award Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>043956-00005B</td>
<td>Docking upAR for Selective Targeting of Cancer Metastasis</td>
<td>New</td>
<td>January 2015 – June 2016</td>
<td>$180,000.00</td>
</tr>
<tr>
<td>044265-00004B</td>
<td>The Bipolar Genome Study</td>
<td>New</td>
<td>May 2014 – April 2016</td>
<td>$87,728.00</td>
</tr>
<tr>
<td>053198-00014B</td>
<td>Collaborative Study on the Genetics of Alcoholism (COGA)</td>
<td>New</td>
<td>September 2015 – August 2016</td>
<td>$1,480,906.00</td>
</tr>
<tr>
<td>054541-00005B</td>
<td>Regulation of RNA Polymerase II Transcription by the Phosphatase Rtr 1</td>
<td>New</td>
<td>August 2015 – July 2016</td>
<td>$291,461.00</td>
</tr>
<tr>
<td>054541-00007B</td>
<td>Regulation of RNA Polymerase II transcription by the phosphatase Rtr 1</td>
<td>New</td>
<td>August 2015 – July 2016</td>
<td>$347,010.00</td>
</tr>
<tr>
<td>056553-00006B</td>
<td>Non-Homologous End Joining Repair in Human</td>
<td>New</td>
<td>April 2015 – March 2016</td>
<td>$219,550.00</td>
</tr>
<tr>
<td>056564-00005B</td>
<td>Metabolic stress responses and Elf2 kinase GCN2</td>
<td>New</td>
<td>May 2014 – April 2016</td>
<td>$343,929.00</td>
</tr>
<tr>
<td>056564-00008B</td>
<td>Metabolic stress responses and Elf2 kinase GCN2</td>
<td>New</td>
<td>May 2014 – May 2016</td>
<td>$116,099.00</td>
</tr>
<tr>
<td>057512-00006B</td>
<td>Early binge drinking and gene regulation</td>
<td>New</td>
<td>September 2015 – August 2016</td>
<td>$225,424.00</td>
</tr>
</tbody>
</table>
2016 IU grant income
$614M

- Non-CI users
  - $148M
  - 24%
- CI users
  - $466M
  - 76%
Grant data capabilities

Current
- IU KFS export -> JSON convert -> XDMoD_VA ingest
- NIH and NSF from public datasets
  - IU to provide scripts for NIH and NSF data imports to local 3 XDMoD_VA instance

Future
- Ability to group by organizational structure (as seen in PTI slide) (IMS export)
Future capability (grouped by organizational structure)
Roles and funding
Grant roles and funding agency

Full XDMoD integration (job data and VA data) if available from your institution.
Grant dollars by funding agency

- NSF
- Lilly Endowment
- NIH-NHLBI
- IU Health
- NIH-NCI
- NIH-NIAID
- DHHS-CMS
- NIH-NIGMS
- S. Dept. of Defense
- NIH-NIDDK
Prorated active grant dollars

Ability to drill down to individual grants
Combined HPC job statistics and value analytics metrics

![Graph showing CPU Hours vs Grant Dollars from Jan 2016 to Nov 2016.](image-url)

- **Jobs:** [CPU Hours: Total]
- **ValueAnalyticsGrants:** [Time-Proportional Dollars of Active Grants ($)]

*Note: The graph shows a comparison between CPU hours and grant dollars over a year, with data points plotted for each month from January to November 2016. The x-axis represents months, while the y-axes represent CPU hours on the left and grant dollars on the right.*
Grants over time associated with funding agencies

Grant size and duration
Publication data
Publication data capabilities

Current
- Working to develop XDMoD schema for ingest
- Access to NIH grant and publication data available
  - NIH Exporter (https://exporter.nih.gov/)
  - Limited visualization capabilities

Future
- Roadmap to link publication and grant data
- Ability to group by organizational structure
- Visualization capabilities integrated
XDMOD Value Analytics

Value Analytics

Funding and Publication Impact
Impact of IT resources on external funding and publications.

Funding Overview
Funding duration, amounts, and types over time.

Co-PI Collaboration Network
Co-PI collaboration network based on NSF funding data.
Integrated view: CI resources, funding, and publications
Visualization: Temporal Bar Graph
Project: XDMoD

This temporal bar graph represents each record as a horizontal bar with a specific start and end year. The width of each bar encodes the total award amount. Bars are colored to represent the funding agency (legend of funding agencies are listed on the right). This graph shows funding duration, amounts, and types over time.
Co-authorship network

Collaboration mapping for publications and grants
Visualization: Co-PI Network
Project: XDMoD

This network represents collaboration patterns based on NSF funding data. Each node represents a principal investigator (PI or Co-PI). It is labelled by the PI’s name, size coded by the total value of all awards for each PI, and color coded by the number of collaborators. An edge between two PIs denotes that they have a grant together with edge thickness denoting the number of times they collaborated and edge color reflecting their success in terms of total sum of all their joint awards. The listing on right rank orders PIs by total dollar amounts.
XDMoD Roadmap

**February 2017**
Pre-Beta Testing Under Way
Grants and PI information available in XDMoD

**Spring 2017**
Beta tester signup
Publication data available
Aggregation charts available

**July 2017**
Beta Release

**June 2017**
Cross-Referencing Grants and HPC Usage Data

**October 2017**
Final Release
Roadmap

Pre-beta testing under way
Grant and PI information available in XDMoD.

February 2017
Beta tester signup
Publication data available.
Aggregation charts available.

Spring 2017
Cross-referencing grants and HPC usage data

July 2017
Beta release

June 2017

October 2017
Final release

Email xdmodva@indiana.edu to participate.
Beta program requirements

- Install and maintain Open XDMoD local$^4$ instance
- Ability to ingest job logs from computational systems
- Ability to get your grant data from your Office of Research Administration
  - We will help you write a script to convert your data to JSON format for ingestion
- Engagement from your institution (you, ORA)
- Rice and SDSC are our two beta testers so far
- Several institutions expressed interest – we’ll be in touch
Thank you.⁵ Questions?

Contact: xdmovda@indiana.edu
License terms

- Please cite as: “XDMoD Value Analytics Module”
- 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 2013 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.