



Distributed
Big Data &
Analytics

Internet2 CINO Distributed Big Data & Analytics (DBDA) Working Group Collaborative Innovation Community Meeting

15 December 2015

Chairs
Alex Feltus, Clemson
Sam Gustman, USC
Marc Hoit, NC State



INTERNET²



POWERED BY
COMMUNITY



Meeting Objectives

- TechEx DBDA Innovation Working Group and Big Idea Hackathon Meeting Update
- DBDA Top Focus Areas
- Engaging with NSF Big Data Regional Innovation Hubs – handout
- Next steps

INTERNET[®]
2

POWERED BY
COMMUNITY



TechEx DBDA WG output: Where Internet2 and Members Can Develop New Innovations

- Agreement on Top 3 focus areas: **Genomics, Smart Cities/Campuses, Digital Humanities**
- Pain points to be addressed:
 - Data management: who owns it & defines it, normalizing it; lack of data schemas
 - Lack of standards on the new ways of collecting data: preservation & exportation
 - Movement of big data environments: more collaboration (norm) “data movement”
 - Moving the data: transparent E2E performance tuning; good network engineers are key
 - Collaboration between various universities on the same data set as it changes
 - Improve the experience for the researchers
 - Improve Researcher/IT/HPC collaboration: Governance models, communication to build trust
 - End to end performance tuning
 - How the DBDA group can leverage NSF Big Data Regional Innovation Hubs: Establishing Spokes to Advance Big Data Applications (BD Spokes)



DBDA Top Focus Areas and updated action plans

Focus Area	Description	Plan
Genomics	Confluence of data from genomics, clinical research, clinical care for precision medicine; Includes medical and agricultural genomics, bioinformatics	<ul style="list-style-type: none"> Align with and leverage NSF Big Data Hubs and spokes efforts Develop HCLS task force with Genomics focus, gather input from NIH and genomics users regarding data, analytics, and network needs and potential innovations
Smart Cities	Sharing, security, privacy, and analytics of data generated in a Smart Grid environment, leveraging the Internet2 network as a testbed, to extend to Smart Cities	<ul style="list-style-type: none"> Align with and leverage NSF Big Data Hubs and spokes efforts Align and collaborate with the IoT innovation working group's Smart Grid and Smart Cities efforts
Digital Humanities	Video/image libraries (e.g., Shoah Foundation), Global sharing of data and content	<ul style="list-style-type: none"> Align with and leverage NSF Big Data Hubs and spokes efforts Potentially develop workshop to increase understanding of data, analytics, network needs, and potential innovations



Regional Networks and Internet2 are Engaging with the NSF Big Data Regional Innovation Hubs and Spokes

- Together, Internet2 and regional networks can provide network and collaboration support for regional BDhubs and spokes
 - Provide private, secure networks for collaboration and testbeds
 - Potential to provide experimental AL2S connectivity for testbeds with multiple participants
 - Potential for setting up “research waves” on the Internet2 backbone (rs@internet2.edu)
 - Provide international networking and data sharing support
 - Enable and facilitate collaborative workshops for intra- and cross-regional BDhubs and “spokes,” for universities, industry, agencies, and labs in domains such as Smart Cities, Health, Genomics
 - Participate in the planning grants for future BDhub spoke proposals focused on collaboration and partnership
 - Support for technology adoption through Operating Innovative Networks workshop series
 - perfSonar for performance optimization

INTERNET²

POWERED BY
COMMUNITY



Regional Big Data Hub Kickoff Meetings and DBDA WG attendees – let us know if you are participating

NSF site: http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505264&org=NSF&sel_org=NSF&from=fund

- **West:** <http://westbigdatahub.org/> - Kickoff meeting held with overall launch 11/3-5 in Washington DC (presentation: <https://vimeo.com/144954304>)
 - Coordinating Institutions: University of California-San Diego, University of Washington, University of California-Berkeley
 - Point of Contact: WesternBDHub@SDSC.edu
- **Midwest:** <http://midwestbigdatahub.org/> - Kickoff webinar 11/24 (presentation: <http://bit.ly/1YbWqr>)
 - Coordinating Institution: University of Illinois at Urbana-Champaign
 - Point of Contact: bdhub-info@ncsa.illinois.edu
- **South:** <http://www.southbdhub.org/index.html> - Kickoff 12/7-8 (presentation: <http://bit.ly/1NJh47f>) - Alex Feltus attended
 - Coordinating Institutions: Georgia Institute of Technology, University of North Carolina at Chapel Hill
 - Points of Contact: aluru@cc.gatech.edu, ashok@renci.org
- **Northeast:** <http://northeastbdhub.dsi.columbia.edu/index.html> - Kickoff 12/16 in NYC - Hudson, Nichols, McMullen
 - Coordinating Institution: Columbia University
 - Point of Contact: NortheastBDHub@columbia.edu

Letters of Intent are due to NSF by January 12, 2016

Proposals with Letter of Collaboration are due to NSF by February 25, 2016

INTERNET[®]

POWERED BY
COMMUNITY



DBDA Working Group Next Steps

- **Establishing a Smart Campus Focus Group**
 - Launching January 2016
 - Explore activities within the Internet2 community
 - Identify and share best practices
 - Develop potential testbed opportunities
 - Develop cohorts by use case, e.g. smart grid, smarter buildings
 - Will be a cross-working group collaboration: IoT, E2ET&S, DBDA
 - Inaugural members include: Arizona State University, Princeton, University of Pittsburgh, University of Washington, University of Wisconsin-Madison
 - Would you like to participate?
- **Healthcare & Life Sciences: Genomics**
 - Internal Internet2 task force assigned to determine a HCLS strategy
 - Hypothesis: many genomics researchers cannot transport their data in a timely fashion and the Internet2 community could enable solutions for bulk data transport obstacles

INTERNET²

POWERED BY
COMMUNITY



DBDA Working Group Next Steps

- **Leverage data flow analysis to engage researchers for collaboration and speed up opportunities**
 - Leverage data flow tools on campus and at Internet2 (Deepfield) to identify data flows
 - Look at data flows to identify researchers using same data, perhaps they can work together, and identify data network leverage opportunities
- **Increase and enable Researcher Engagement to understand and support DBDA needs**
 - Leverage NSF “[Broadening The Reach](#)” model and cookbook (cookbook TBA)
 - Leverage Internet2, ESnet, IU [Operating Innovative Network \(OIN\) workshops](#) on networking capabilities
 - Use Internet2/ESnet/GEANT [SCBOF survey](#) re researcher engagement for HPC/HTC/Networks
 - Cross Connects Workshop Series: collaboration Internet2 & ESnet
- **Potential workshops on collaboration/partnering around campus data management**
 - Better understand data, analytics, and network needs
 - By use case areas – Potential workshops for HCLS/Genomics and separately Digital Humanities
 - Attendees include IT that supports research and researchers
 - Cultural aspects: See “[CIOs reach across the aisle to enable collaborative research innovation](#)” EDUCAUSE presentation, can use questions in the back-up on campus in workshops

INTERNET²

POWERED BY
COMMUNITY



Innovation Working Group Operations Timeline: 2015 – Startup year



1Q

- Member survey on open collaborative innovation ideas

2Q

- Announce working groups based on member survey
- Attract co-chairs and members
- Working Groups Startup
- Monthly meetings
- Gather use cases

3Q

- Cross-working group discussions
- Develop plans: tactical & strategic

Tactical outcomes

- E2ET&S SDP Webinar
- IoT Sandbox

4Q

- TechEx in-person meetings to discuss recommendations
- Gather new input

Strategic Plans

- Smart grid testbed
- E2ET&S for IoT
- Healthcare/Life Sciences



Innovation Working Group Operations Timeline: 2016 We evolve and grow

1H

- Develop Healthcare/Life Sciences (HCLS) strategy with DBDA
- Develop DBDA Digital Humanities/Humanists strategy
- Smart Campus focus group
- E2ET&S for IoT workshop(s)
- Smart Grid testbed planning
- Expand IoT Sandbox
- Increase researcher engagement
- Continue education & awareness for new innovation opportunities

2H

- HCLS strategy execution
- DBDA Digital Humanities strategy continuation
- Continue IoT and DBDA strategy development
- E2ET&S collaboration across extended community
- Testbed pilots – Smart Grid/Campus, HCLS
- Smart Campus enablement
- Assess/extend Sandbox approach
- Constant market insight gathering
- Innovation input from community

Ongoing Community Input

INTERNET[®]
2

POWERED BY
COMMUNITY



Next Steps

- Leverage NSF Big Data Regional Innovation Hubs to address DBDA pain points and build community across hubs
- Working with regional networks to create joint value proposition for Big Data Regional Innovation Hubs and spokes
 - Let us know if you are participating in the BDhubs
 - Leverage the joint value prop on page 5, feedback welcome
- Kicking off smart campus/cities focus group 1Q16
 - Let us know if you want to participate
 - Share best practices and create new practices
 - Leverage Internet2 and regional networks as testbeds





Questions & Answers...
Thank You
dbda-chairs@Internet2.edu
CINO@Internet2.edu
[@FIoInternet2](#)
[@EmilyNInternet2](#)

INTERNET²

**POWERED BY
COMMUNITY**

