

# Cyberinfrastructure Collaboration in the Northeast

John Goodhue  
Executive Director  
MGHPCC  
October 3, 2017

# Topics

- One region, many lenses
- Some collaborative works in progress
- Observations

# Different Lenses Serve Different Purposes

Northeast Cyberteam    ME NH VT MA

XSEDE Region 7        ME NH VT MA CT RI NY

Northeast Big Data Hub   ME NH VT MA CT RI NY NJ PA

# Even the Parochial Lenses



# The Cyberinfrastructure Stack

## CI Engineering Roles

## Platforms

## Accountability

Research Computing  
Facilitator

simulation, instruments, machine  
learning, performance tuning,  
workflows, data analytics, data  
management, data access, data  
management and sharing...

Research

System Administrator

servers, storage, file systems,  
object stores, cloud integration,  
performance tuning, data centers,  
containers, authentication,  
schedulers...

and

Network Engineer

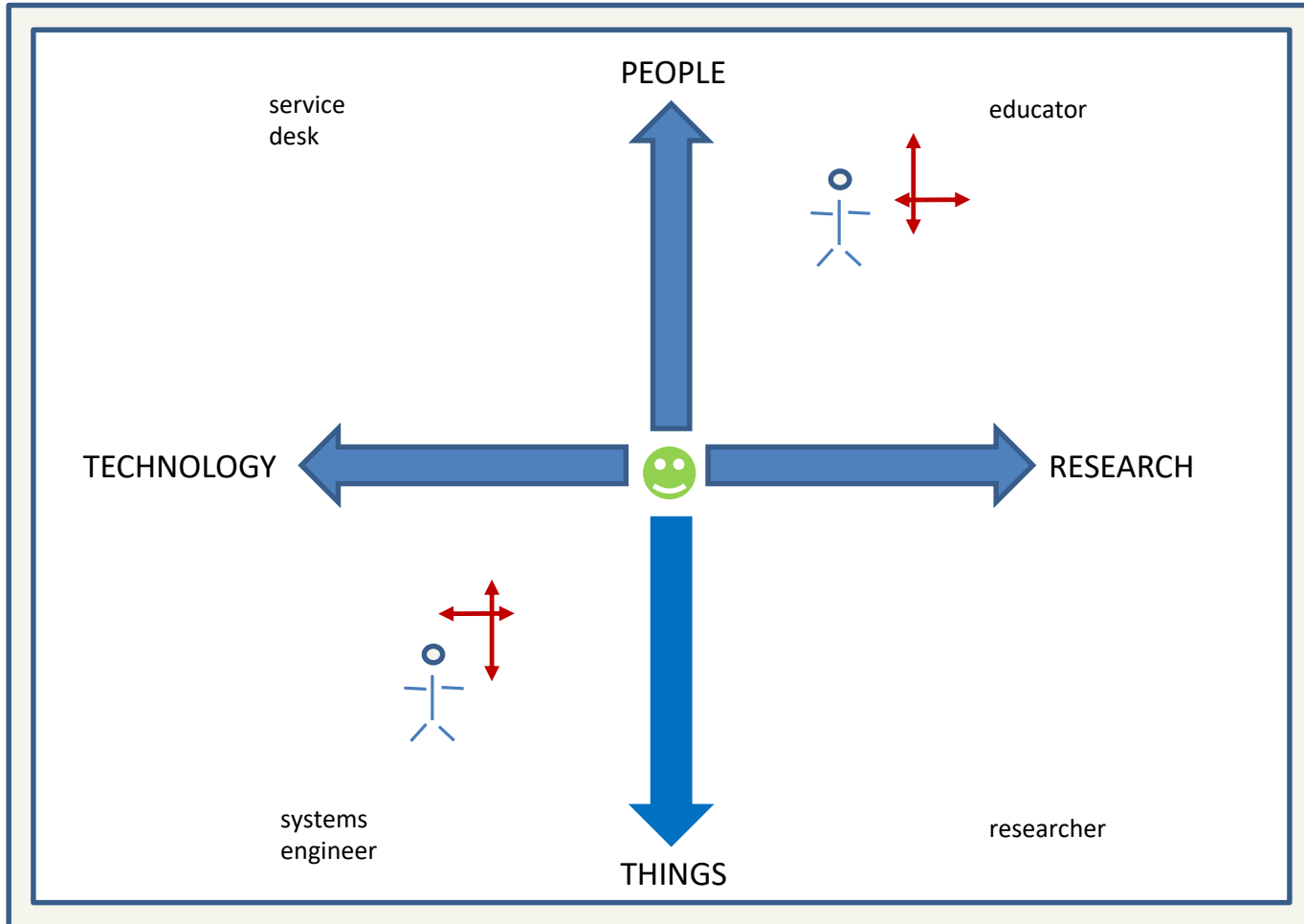
Links; service provider contracts;  
Peering; Firewalls; bottlenecks; DDOS  
mitigation; Science DMZs; intrusion  
detection; manhole fires; SDN...

Information  
Technology

In science if you know what you are doing you should not be doing it.  
In engineering if you do not know what you are doing you should not be doing it.

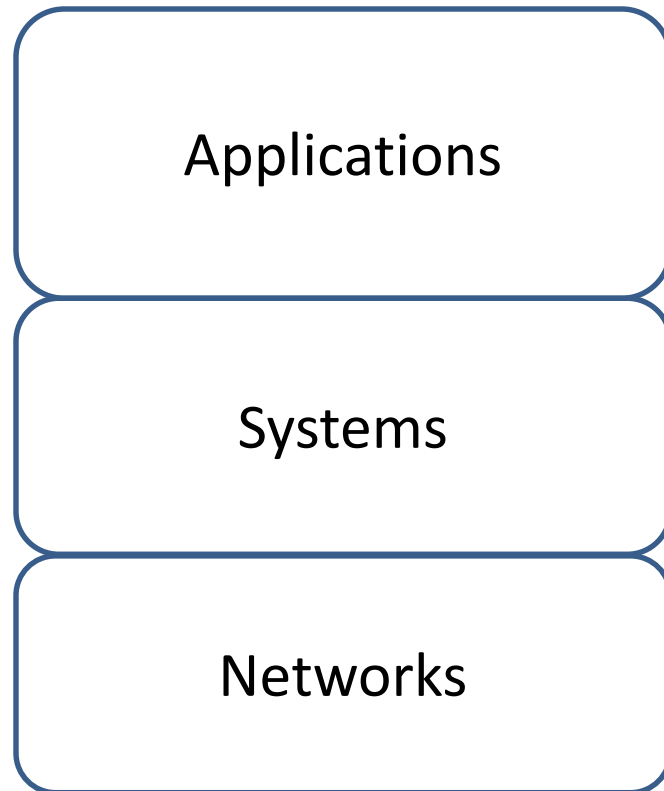
Richard Hamming

# The Cyberinfrastructure Spectrum



Neeman/Cuff 2016

# Collaborative works in Progress



Region 7 XSEDE champions  
ACI-REF  
Northeast Big Data Innovation Hub  
Northeast Cyberteam

MGHPCC  
Engaging-1  
C3DDB  
Mass Open Cloud  
Northeast Storage Exchange

Northeast Research and Education Network  
Northern Crossroads (NoX)  
New Hampshire science DMZ

# Case Study – Northeast Cyberteam

- Started May 2017
- Focus
  - Research Computing Facilitators at small research institutions in NNE
- Problem
  - Not enough of them; and they are hard to retain
- Approach
  - Apply student/mentor teams to immediate research problems
  - Leave tools and knowledge in our wake
  - Grow a sustainable shared pool of Research Computing Facilitators



# Origins



**John Goodhue**  
Executive Director  
Massachusetts Green High  
Performance Computing Center



**Stephen Everse**  
Associate Professor  
University of Vermont



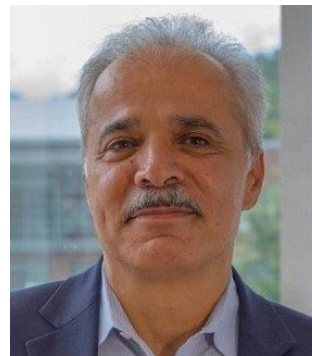
**Bruce Segee**  
Professor  
University of Maine



**Scott Valcourt**  
Director of Strategic Technology  
University of New Hampshire



**Julie Ma**  
Regional RCF  
Massachusetts Green High  
Performance Computing Center



**Sia Najafi**  
Interim CIO  
Worcester Polytechnic  
Institute

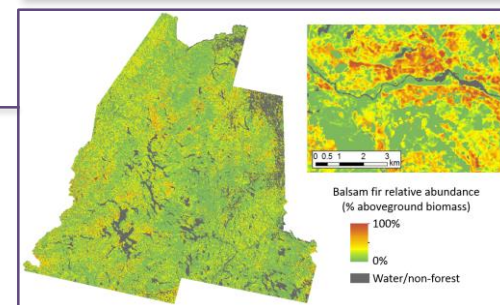
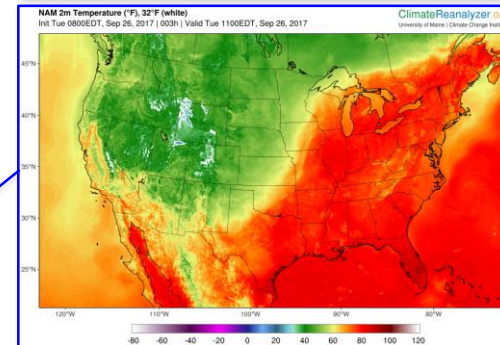
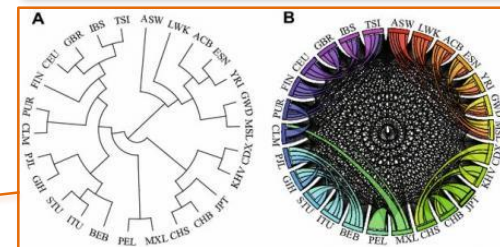
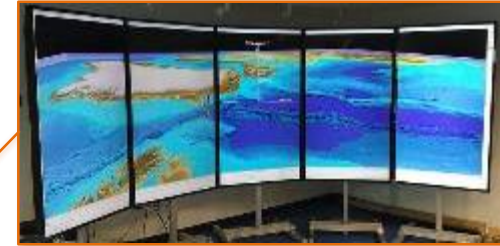


**Ralph Zottola**  
CTO-Research Computing  
University of Massachusetts

# Step 1 – Build a Project Pipeline

## Projects

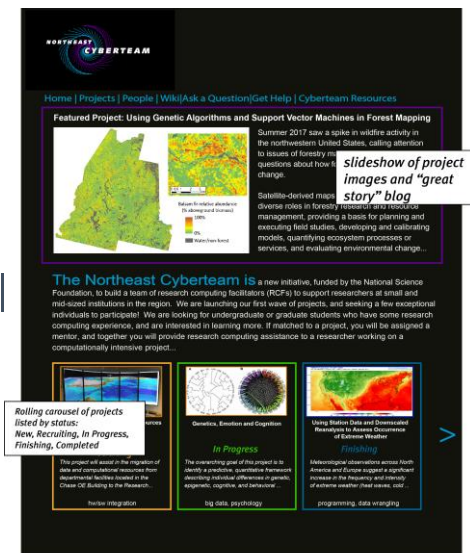
Project Title	Student Profile	Target Start Date	Location
AR Mapping of Historic Higgins House	Graduate CS student to supply HPC expertise to help offload mapping and image compositions.	January, 2018	Becker College
Understanding, applying and documenting multi-core and Spark techniques	Graduate student in any area of Computational Science and Engineering; strong writing skills	October, 2017	Bentley University
Centralizing Resources for Sea Floor Mapping	Undergraduates with an interest in system design and programming HPC devices. A lack of a fear of hardware and software is most desired	July, 2017	University of New Hampshire
Genetics, Cognition and Emotion	A graduate student with expertise in dealing with large data sets	July, 2017	University of Vermont
Developing a Regional HPC Help Desk	Graduate student/Post Doc in any area of Computational Science and Engineering, with knowledge of programming and applications	August, 2017	Worcester Polytechnic Institute
Developing Computational Labs for an Upper Level Physical Chemistry Course	An undergraduate/graduate student with chemistry focus that has taken physical chemistry courses; has knowledge of computational modeling and experience with Gaussian	September, 2017	Bridgewater State University
Using Station Data and Down-scaled Reanalysis to Assess the Occurrence of Extreme Weather	An undergraduate student with a knowledge of programming and data handling	July, 2017	University of Maine
Using Genetic Algorithms & Support Vector Machines in Forest Mapping	An undergraduate student familiar with Matlab, C, and High Performance Computing	July, 2017	University of Maine



New and recruiting
  Recruiting
  In Process
  Finishing up
  Complete

# Next Steps

- Expand the project pipeline / solidify intake and matching process
  - Aiming for 40+ projects over 30+ institutions over 3 years
- First wave of "tools in our wake"
  - Build awareness of existing resources (ACI-REF, XSEDE training, et al)
  - StackExchange for Q&A
  - Regional help desk
- Start work on the foundation for a shared regional pool
  - Project pipeline to demonstrate value
  - Organizational underpinnings for resource sharing



# Other Opportunities and Challenges

- Security
- The Long Tail of science
- New devices: GPU, FPGA, nVME, crosspoint
- Data explosion / data retention and curation
- Cloud services / vendor relationships
- Operational convergence
- Outreach / broadening participation
- Changing operating system landscape
- Algorithmic complexity

# Observations

- Researchers routinely collaborate; institutions not so much
- Things that help with institutional collaboration
  - A little bit of senior administration attention (President, VPR, CIO)
  - Trust, respect, and personal relationships (hard to build; easy to break)
  - Rules of engagement (enforce simplicity, fairness, transparency, predictability)
  - OSP, OGC, internal audit et al are your friend, really
  - Grounding infrastructure projects with science and education objectives
  - Focus on long term results
- Time is the principal bottleneck
  - People don't scale very well
  - Automating the easy stuff is essential
  - Looking beyond the traditional sources is an under-explored opportunity

