Mass Big Data:
Progressive Growth through Strategic Collaboration

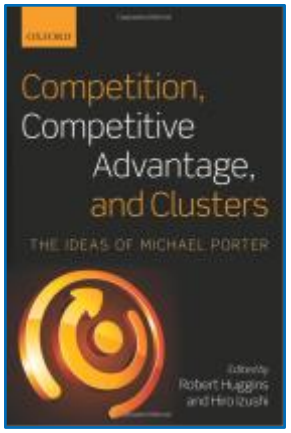
Patrick Larkin, *Executive Director*
The Innovation Institute at the
Massachusetts Technology Collaborative

October 15, 2014

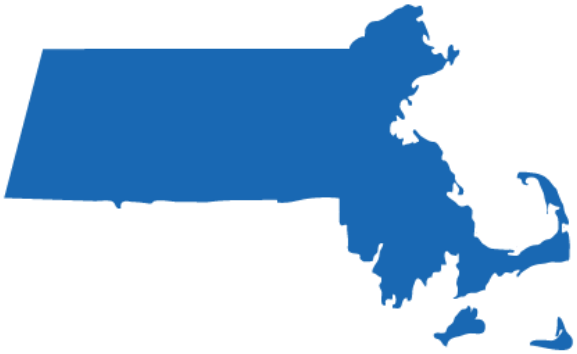
Partnering with the Innovation Institute

Established by MA State Legislature in 2003, **The Innovation Institute at The Massachusetts Technology Collaborative** provides thought leadership and organizing concepts to sustain and grow the innovation economy, with implications for public policy.

Cluster Growth



Regional Growth



Matching Fund

Industry/Economic Analysis

State Investments

Invest **$62M**

Leverage **$347M**

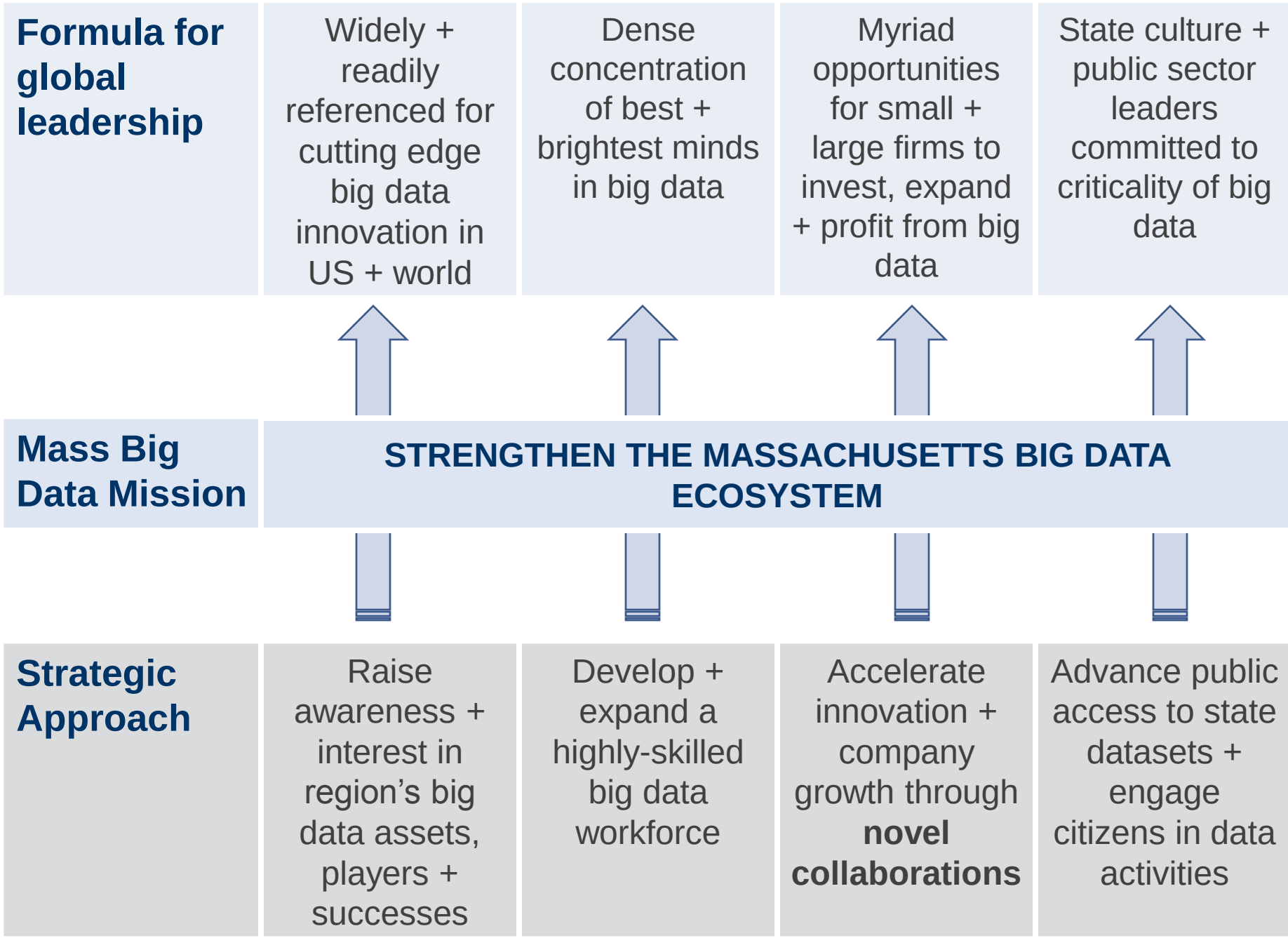
High Value Projects

The Mass Big Data Initiative

Formula for Global Leadership

Mass Big Data Initiative
launched by Governor Patrick in 2012

- **500** Companies
- **10** leading research centers
- **5,600** graduates from **14** data
- science-related programs
- **$2.5B** investment in **Big Data**
- **Highest Density** of **Big Data** talent in the US is in **MA**.


Formula for global leadership
Widely + readily referenced for cutting edge big data innovation in US + world
Dense concentration of best + brightest minds in big data
Myriad opportunities for small + large firms to invest, expand + profit from big data
State culture + public sector leaders committed to criticality of big data
Mass Big Data Mission
STRENGTHEN THE MASSACHUSETTS BIG DATA ECOSYSTEM
Strategic Approach
Raise awareness + interest in region's big data assets, players + successes
Develop + expand a highly-skilled big data workforce
Accelerate innovation + company growth through novel collaborations
Advance public access to state datasets + engage citizens in data activities

MASS BIG DATA SIGNATURE PARTNERSHIPS

Industry- Academia - Public Sector

- Mass Green High Performance Computing Center
- Mass Open Data Repository
- Massachusetts Open Cloud
- MassDOT data analysis and application development
- Mass Big Data Tech Treks
- Mass EduData Projects
- hack/reduce Ambassador Program

MASS OPEN CLOUD

An Open Computing Marketplace …
Negotiating Whatever Services
you need from Multiple Vendors

3 year project to build public cloud that will serve as a shared infrastructure in the Commonwealth for running computationally intensive and Big Data applications.

MASS OPEN CLOUD…CONT'D

- BU's Center for Cloud Innovation partnering with MGHPCC, its four partner universities, tech companies, and Oak Ridge National Lab
- Designed to embrace software, hardware, and services companies as partners in a multi-sided marketplace
- All stakeholders have access to operational data and cloud operations
- OCX (Open Cloud eXchange) system creates an open economic model where partners decide how to charge for their services
- OCX design reduces the responsibility of the provider and broadens the availability of performance and operational data
- MOC ecosystem created to generate economic development and innovation by engaging partners in the marketplace

MOC VALUE

Benefits to Commonwealth Users:

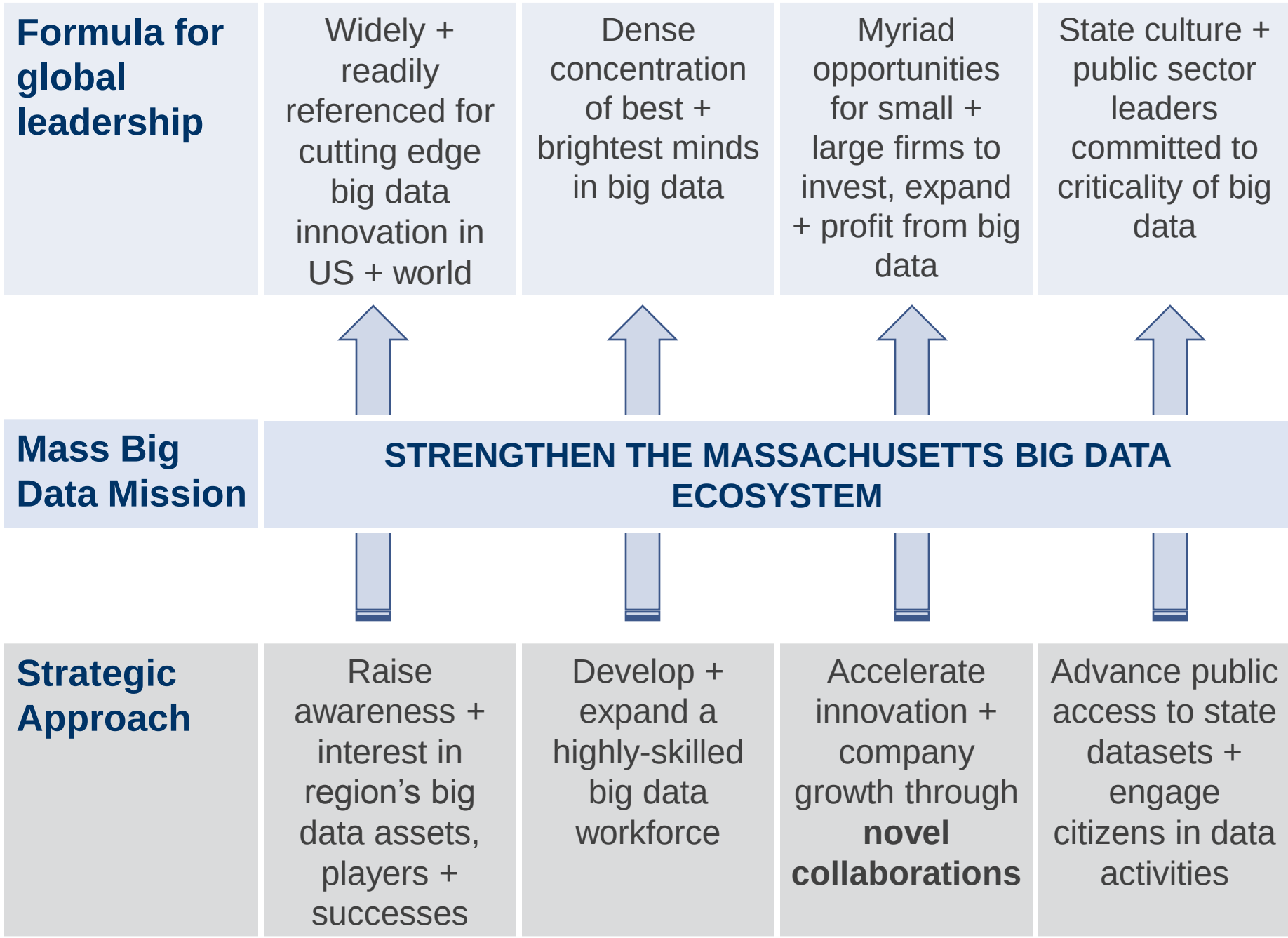
- Can move data into, within, and out of the cloud efficiently and inexpensively
- Can evaluate and use services from many vendors
- Access to large public data sets, including the public data sets of the Commonwealth.

Benefits to Industry

- Can integrate, test and certify HW and SW
- Demonstrate technologies for private and public clouds
- Evaluate (new) products with real customers at an early stage
- Engage with researchers and students; workforce development

Benefits to Academic Institutions:

- A unique platform for research on cloud computing
- Increased competitiveness for funding from federal agencies, foundations industrial partners.


Formula for global leadership
Widely + readily referenced for cutting edge big data innovation in US + world
Dense concentration of best + brightest minds in big data
Myriad opportunities for small + large firms to invest, expand + profit from big data
State culture + public sector leaders committed to criticality of big data
Mass Big Data Mission
STRENGTHEN THE MASSACHUSETTS BIG DATA ECOSYSTEM
Strategic Approach
Raise awareness + interest in region's big data assets, players + successes
Develop + expand a highly-skilled big data workforce
Accelerate innovation + company growth through novel collaborations
Advance public access to state datasets + engage citizens in data activities

MOHPCC

MASSACHUSETTS GREEN HIGH PERFORMANCE COMPUTING CENTER

Massachusetts Green High Performance Computing Center

National Academies GUIRR

October 15, 2014

BOSTON UNIVERSITY · MIT · Harvard · UMASS · Northeastern · CISCO · EMC²

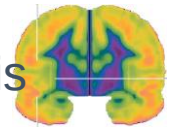
The MGHPCC Data Center and Consortium

A partnership between 5 universities….

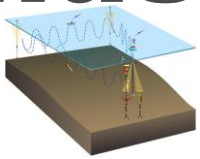
The Commonwealth, and industrial sponsors

Data-intensive Research in Massachusetts

Medical imaging analysis



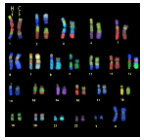
(BU, Harvard, Children's Hospital)



Instrumenting the ocean floor (UMass, WHOI, MIT)

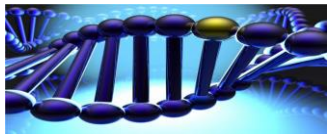
SCOPE
Smart-city Cloud-based Open Platform and Ecosystem (BU)

Understanding Evolution (UMass)

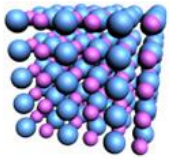


ATLAS/LHC (BU, Harvard)

Bicep2 / Gravity Waves (Harvard)



DNA samples -> Treatment Strategies (UMass, MIT, Harvard Medical)

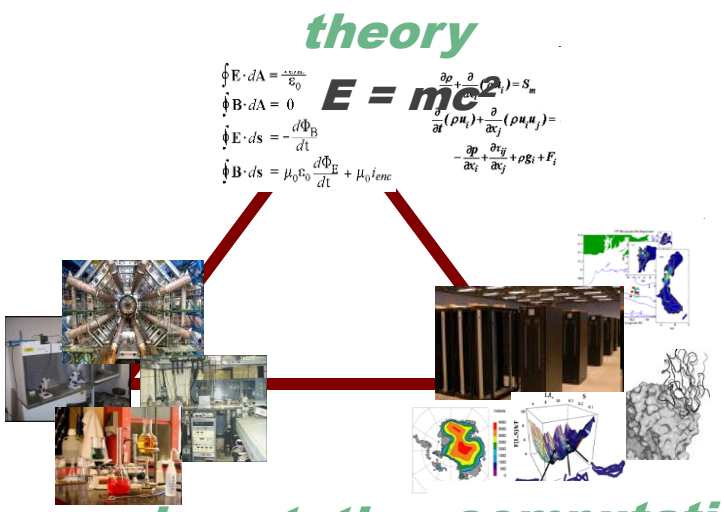


Computational Nanotechnology (Northeastern, Harvard)

Three Works in Progress

- Regional data center dedicated to research computing
- University/Industry/Government research collaboration
- Regional economic development

Motivation



Computing is now the “third Leg” of Science and Engineering”

2008

2012

Incremental expansion of computing infrastructure no longer sufficient

MGHPCC Data Center

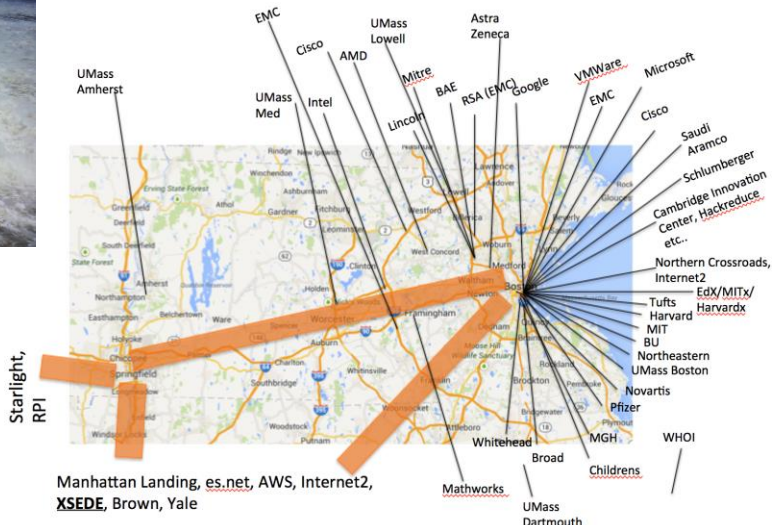
10 MW for compute / 15MW total

8 Acres and a 40MVA power feed for expansion

Green, low cost power

LEED platinum certification

Communication Crossroads



Research and Operational Collaboration

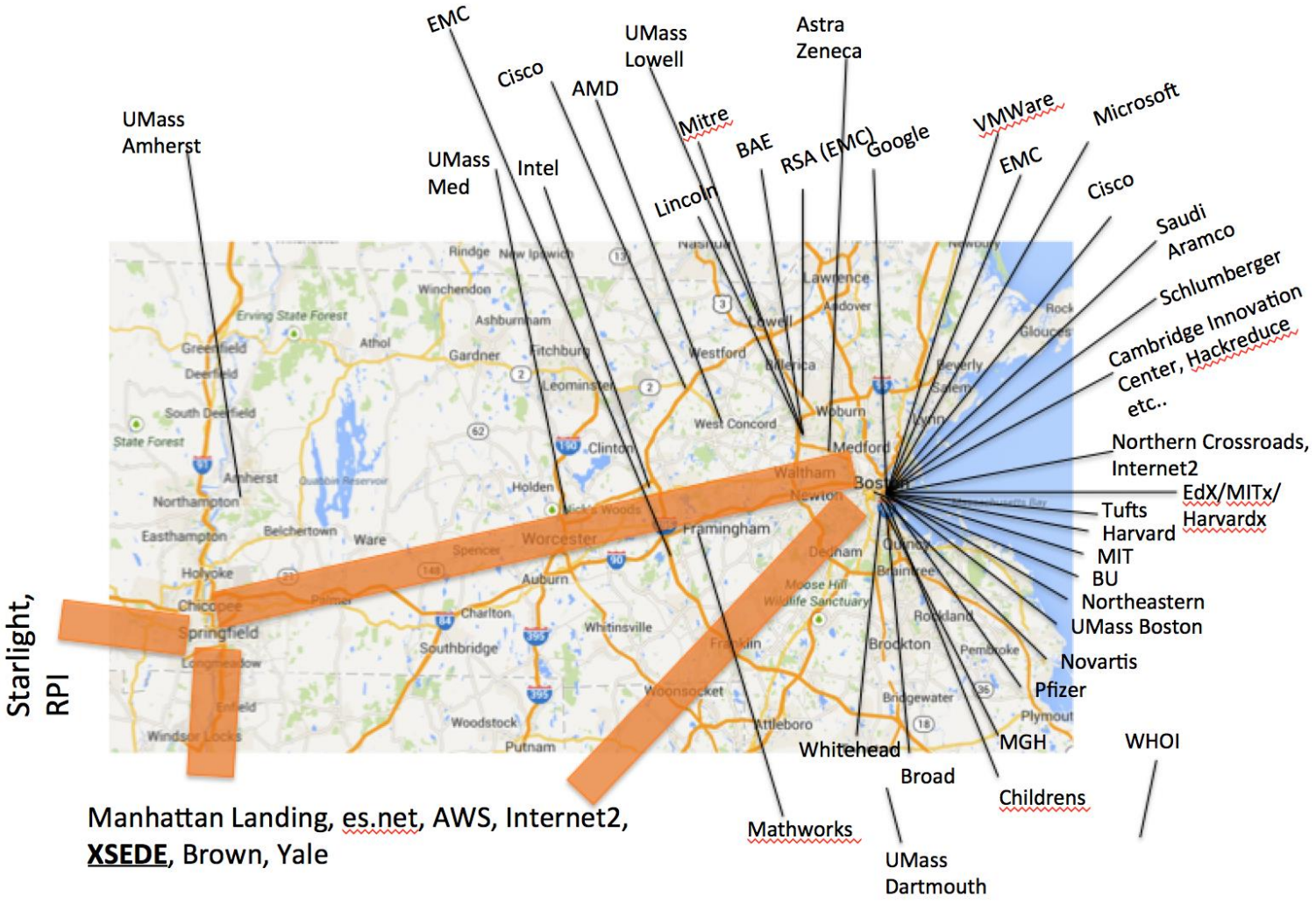
Seed Grant Program::
- Multi-university collaboration
- Potential to lead to larger initiatives
- $1M over 2 years

engaging1::
- Science via traditional workflows and systems
- Non-traditional workflows and technologies
- Experiments with joint facilities integrated across universities

C3DDB::
- Dedicated to Life Sciences research
- University and industry user community
- Mix of conventional and forward-looking hardware

Mass Open Cloud::
- New approach to cloud computing infrastructure
- More than a dozen university, industry and government collaborators

Economic Impact



The Next Generation

Hands-on Construction Engineering Classroom

Educators Conference

Science in the Cloud

Innovation District Task Force

Cisco Network Academy

Clean Energy Innovation Workshop

Scholarship Fund

Education Opportunities Workshop

Smart Meter Project

Girls Inc Robotics Competition

Closing Thoughts

- What we have done so far
  - **Deliver Operationally**
    2 years successful collaboration on common infrastructure
  - **Show Research Collaboration Proof Point**s
    Pooling resources across institutions can deliver powerful results
  - **Catalyze Near Term and Longer Term Economic Impact**
    Research and education


MGHPCC
MASSACHUSETTS GREEN HIGH PERFORMANCE COMPUTING CENTER
BOSTON UNIVERSITY
UMASS
MIT
HARVARD UNIVERSITY
Northeastern
Bigelow Street
MGHPCC
Park
P