From Genes to Stars

idies

Institute for Data Intensive Engineering and Science

idies

Krieger School of Arts and Sciences,
Whiting School of Engineering,
Sheridan Libraries at Johns Hopkins University,
Johns Hopkins School of Public Health
and the
Johns Hopkins Medical Institutions

Data-Intensive Research at JHU

- Sloan Digital Sky Survey
- VAO/LSST /Pan-STARRS
- Multi-scale materials science
- Earth circulation modeling
- Environmental sensors
- Turbulence, fluid dynamics
- Large Hadron Collider

- High Throughput Genomics
- Biostatistics with Big Data
- Personalized health
- Connectomics
- OncoSpace, Radiology
- BIRN
- BME

Data Conservancy, DIBBs

- JHU investments
- NSF grants
- Keck/Moore grants
- Discovery grants

- GrayWulf (2008, 1PB)
- HHPC* (2009)
- Data-Scope (2011, 10PB)
- BayView HPRCF (2015)

Industry: Microsoft, Google, NVIDIA, Intel, Dell

Some basic statistics

- Lots of new people joined (over 80 faculty)
- Seed funds (9 grants so far)
- Wide ranging spectrum of projects
- Increasing presence in East Baltimore
- New federal grants

pending
declined
funded

Cumulative funding

$M
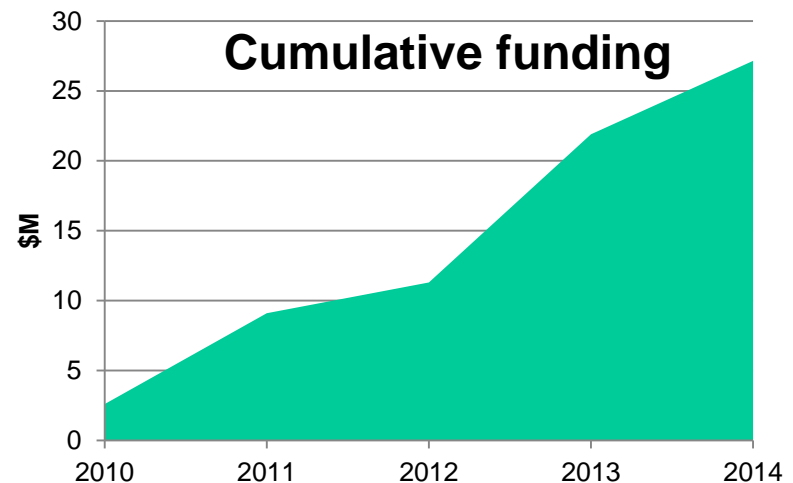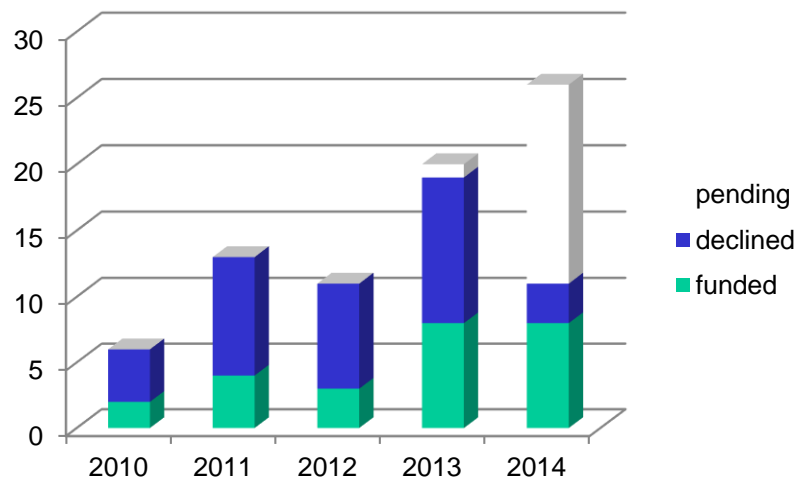
Sociology of Big Data Science

- ## Broad sociological changes
  - *Convergence of Physical and Life Sciences*
  - *Data collection in ever larger collaborations*
  - *Virtual Observatories: CERN, VAO, MGI, NCBI, NEON,…*
  - *Analysis decoupled, off archived data by smaller groups*
  - *Emergence of the citizen/internet scientist*
  - *Humanities and economics are also moving into Big Data*
- ## Need to start training the next generations
  - *Π-shaped vs I-shaped people*
  - *Early involvement in "Computational thinking"*
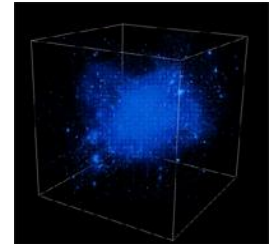
The Mission of IDIES

- **Intellectual leadership**

  in the "Science of Big Data"

- **Incubator**

  for data intensive discoveries

- **Vision and oversight**

  of high performance and data intensive computing

- **Train the next generation**

  in data analytic skills

- **Support unique data resources**

  which give us competitive advantages and visibility

IDIES Resources

- Help and coordination for submission of large interdisciplinary grant proposals
- Enable leveraging of large, unique data resources
  - *Sloan Digital Sky Survey, several 100TB of turbulence simulations, in connectomics, over 2PB in data collections*
  - *These give an "unfair advantage" to JHU scientists*
- Share special data intensive computational resources
  - *Data-Scope, 100G networking to Internet 2 and across JHU*
  - *MARCC: new computational facility shared with UMCP*
- Help in starting new data intensive projects
- Help in bringing (and analyzing) your big data

http://idies.jhu.edu/

idies

Annual Review 2014

Oct. 1, 2013 - Sept. 30,