Making the Web Work for Science: The Impact of e-Science and Cyberinfrastructure

CENDI/NFAIS/FLICC Workshop

December 8, 2008

Chris Greer
Director, US National Coordination Office
Networking and Information Technology Research and Development Program

Acronyms:

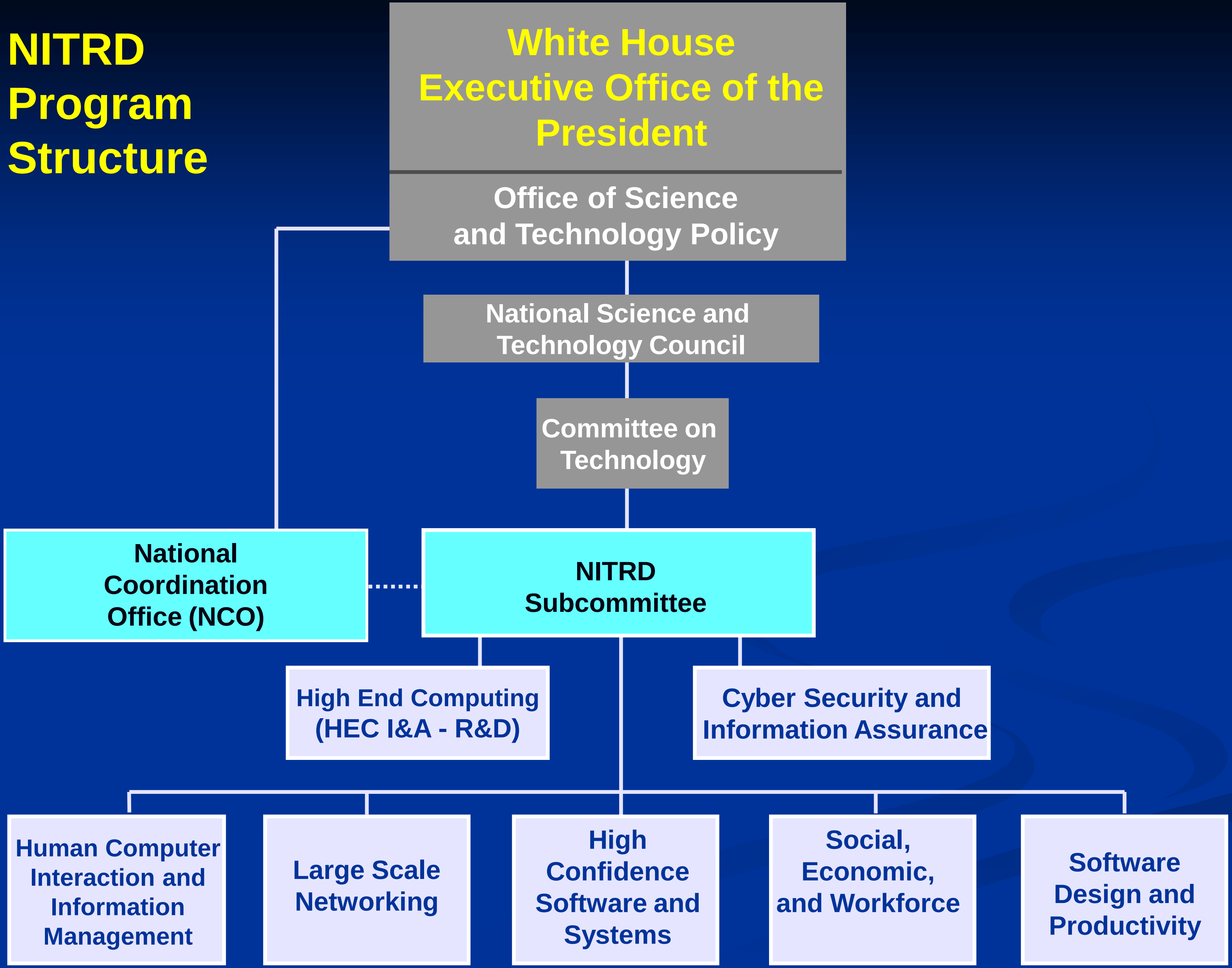
NITRD

Networking and Information Technology Research and Development Program

NCO

National Coordination Office

NITRD Program Structure



AHRQ Agency for Healthcare Research and Quality

DARPA Defense Advanced Research Projects Agency

DOE/NNSA Department of Energy - National Nuclear Security Agency

DOE/SC Department of Energy - Mathematical, Information, and Computational Science Division

EPA Environmental Protection Agency

NARA National Archives and Records Administration

NASA National Aeronautics and Space Administration

NIH National Institutes of Health

NIST National Institute of Standards and Technology

NOAA National Oceanic and Atmospheric Administration

NSA National Security Agency

NSF National Science Foundation

OSD and DoD Service research organizations, Office of the Deputy, Under Secretary of Defense (Science and Technology)

Science is global and thrives in a world that is not limited to 4-dimensions

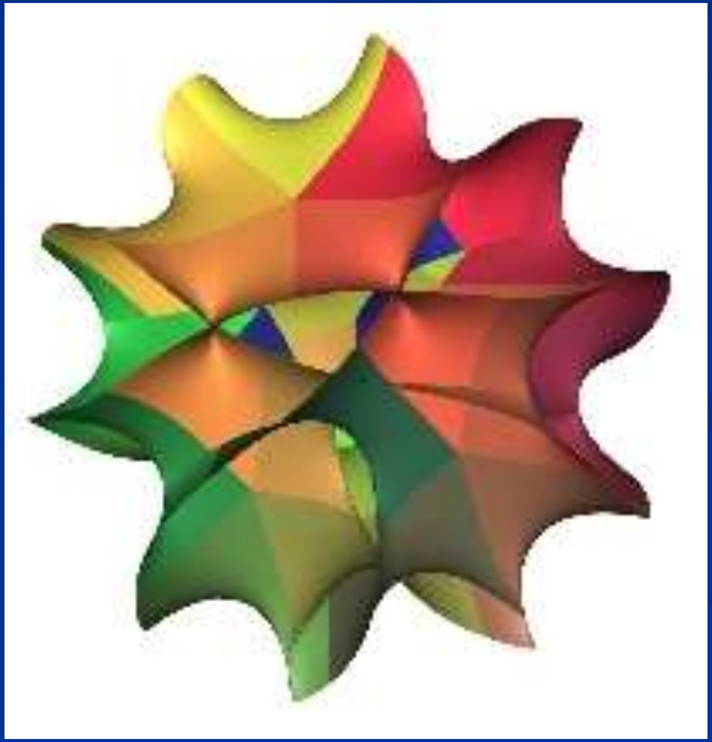
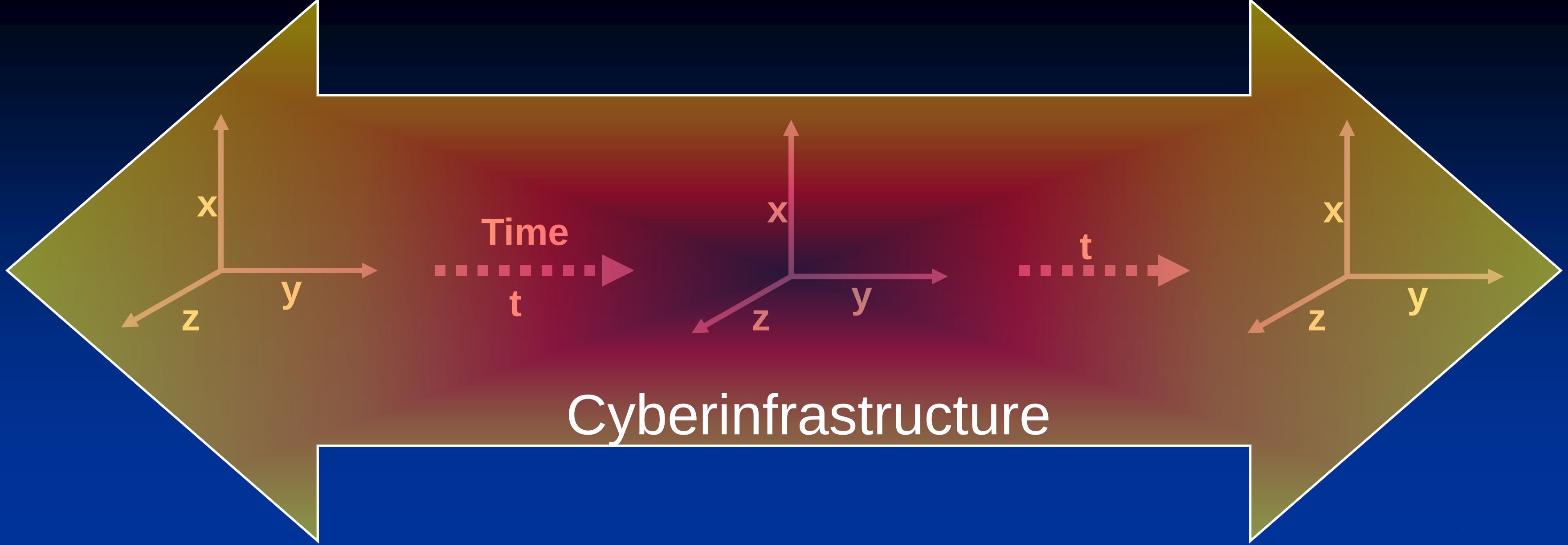


Image: Andrew J. Hanson
www.cs.indiana.edu/~hanson/

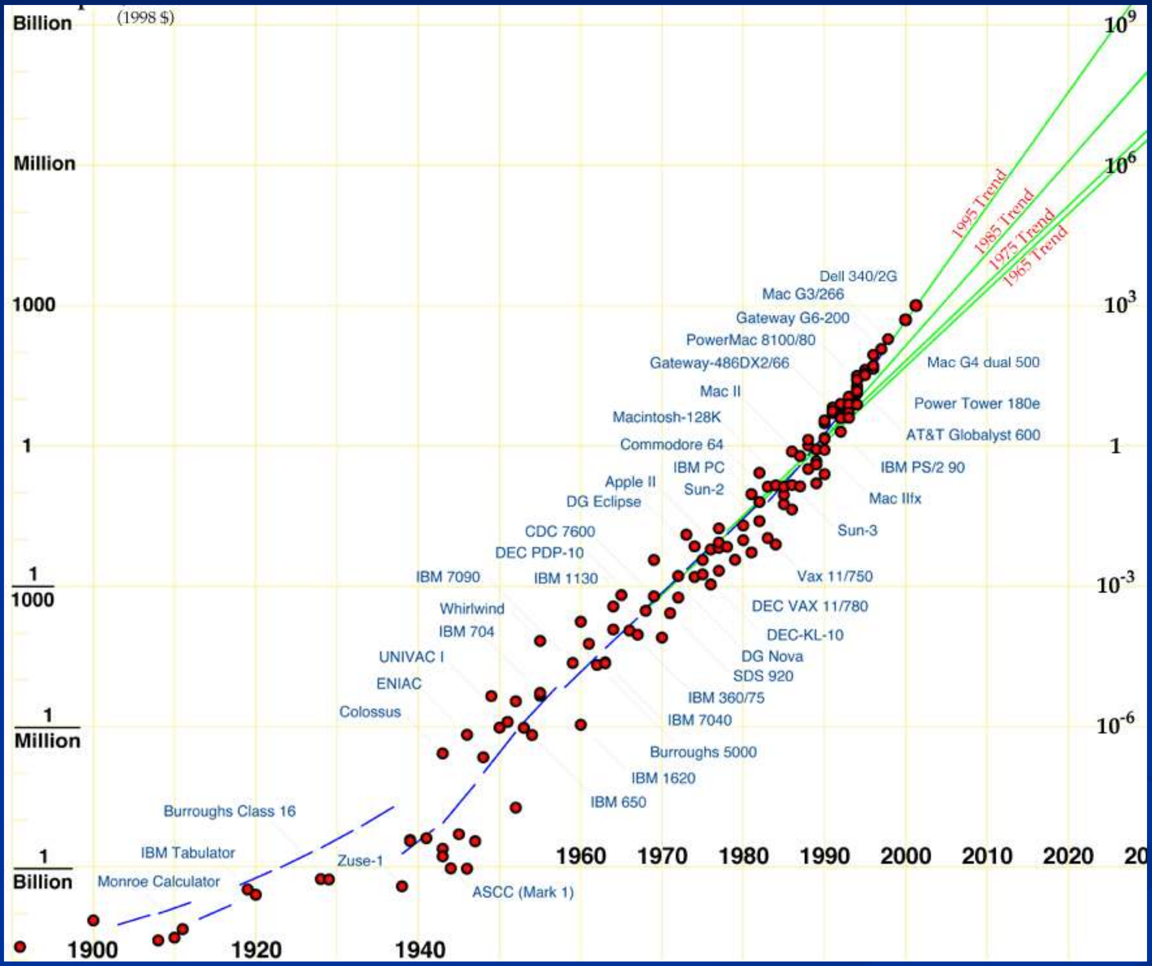


Computational capacity and capability

Information for innovation and discovery

Connectivity for access and interaction

Increasing Computational Capacity and Capability



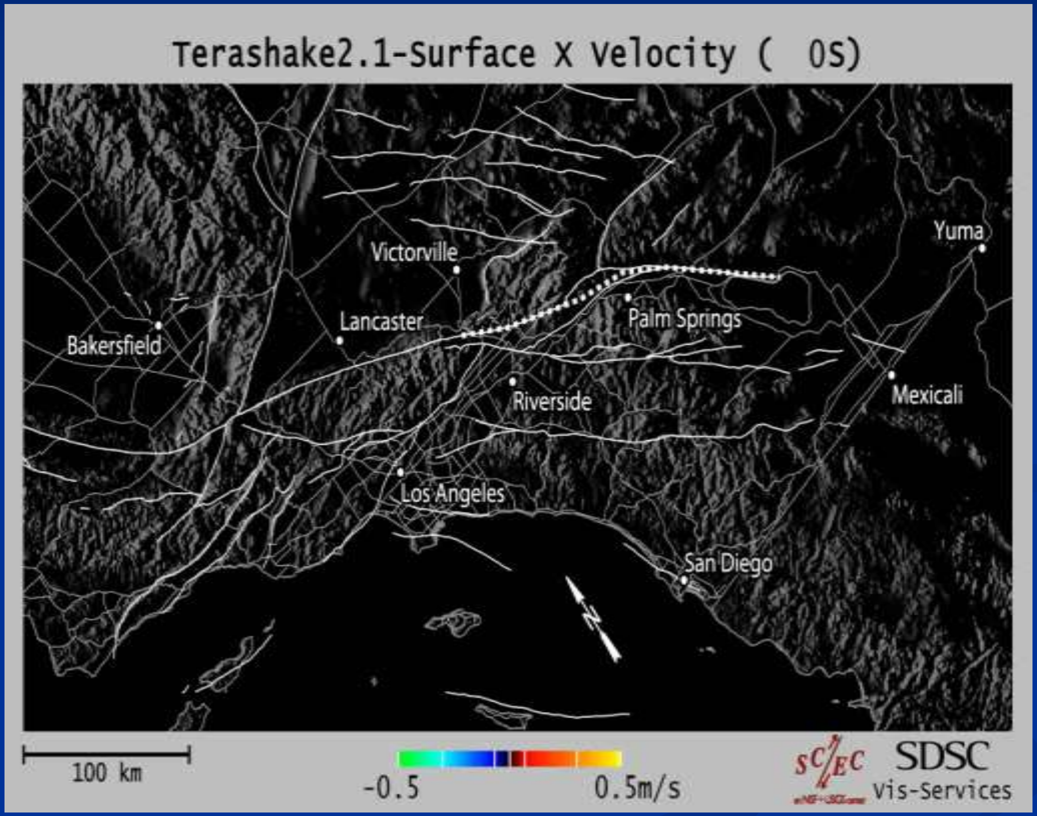
Source: Hans Moravec, "Robots After All,"
Carnegie Mellon University, Aug. 2003

ENIAC, dedicated in 1946, was one of the first fully-functional digital computers, using 17,000 vacuum tubes for up to 5,000 addition operations per second. Today's petascale machines are designed to sustain more than one quadrillion (1,000,000,000,000,000) operations per second. A calculation these machines could complete in a week would take a machine operating at ENIAC speeds several billion years.

TeraShake Simulation

Simulation of 7.7 earthquake on lower San Andreas Fault

- Physics-based dynamic source model – simulation of mesh of 1.8 billion cubes with spatial resolution of 200 m
- Builds on 10 years of data and models from the Southern California Earthquake Center
- Simulated first 3 minutes of a magnitude 7.7 earthquake, 22,728 time steps of 0.011 second each
- Simulation generates 45+ TB data



Source: Fran Berman, Director, SDSC

Behind the Scenes – Enabling Infrastructure for TeraShake

- *Computers and Systems*
  - 80,000 hours on 240 processors of DataStar
  - 256 GB memory p690 used for testing, p655s used for production run, TG used for porting
  - 30 TB Global Parallel file GPFS
  - Run-time 100 MB/s data transfer from GPFS to SAM-QFS
  - 27,000 hours post-processing for high resolution rendering
- *People*
  - 20+ people involved in information technology support
  - 20+ people involved in geoscience modeling and simulation
- *Data Storage*
  - 47 TB archival tape storage on Sun StorEdge SAM-QFS
  - 47 TB backup on High Performance Storage system
  - SRB Collection with 1,000,000 files

Source: Fran Berman, Director, SDSC

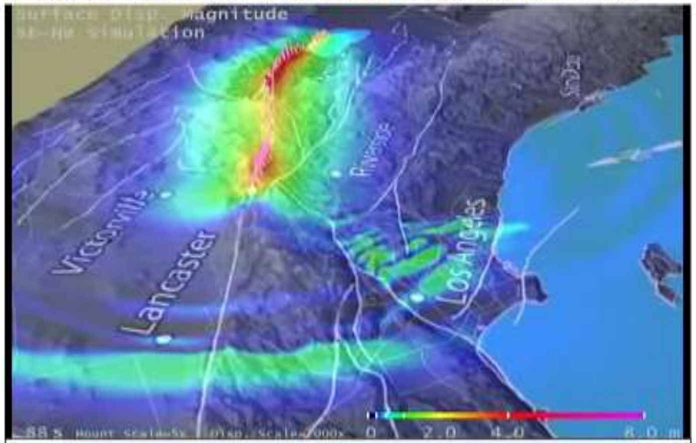
username ●●●●●●●●● Log in Register Request new password Search SciVee Search

SciVee beta — make your research known

browse | upload | community | Help

my communities | communities list | discussions list

TeraShake: A Southern California Earthquake Simulation



0:26 1:25 Menu

▾ Comments

SOALLOW (Feb 12, 2008 01:28)

THANK'SSSSS

You must be logged in to comment

General | Share | Communities | Tags | Viewers' Notes | Ask

Submitted by: sdsc

Description: The TeraShake simulations modeled the earth shaking that would rattle Southern California if a 230 kilometer section of the San Andreas fault ruptured producing a magnitude 7.7 earthquake. Two rupture scenarios were simulated, one rupturing from north to south, beginning near Wrightwood, California, and a second one rupturing from south to north, starting near Bombay Beach, California. To model... » More

Rating: ★★★★☆

You must be logged in to rate a video.

Uploaded: Monday, October 8, 2007

Views: 13816

Comments: 1

Favorites: Add to Favorite

Report Abuse: Flag this video

Supplementary Materials

There aren't any supplementary materials available for this video.

Add a supplementary file

Related videos | More videos from this user

TeraShake: Large-scale Simulations of...
Views: 2997 | Uploaded: Oct 31, 2007
submitted by: sdsc

The Shaky Future of the Salton Sea
Views: 7476 | Uploaded: Jan 8, 2008
submitted by: Scripps Institution of Oceanography

Visualization of Southern California...
Views: 551 | Uploaded: Oct 11, 2008
submitted by: amit_chourasia

Part1: Managing Simulation Output
Views: 8582 | Uploaded: Oct 29, 2007
submitted by: sdsc

Part 2: Managing Simulation Output
Views: 8580 | Uploaded: Oct 29, 2007
submitted by: sdsc

```
SEQADV 3F13 MET A    2  UNP  Q7NWQ3        EXPRESSION TAG
SEQADV 3F13 SER
SEQADV 3F13 GLU
SEQADV 3F13 GLY
SEQADV 3F13 HIS
SEQADV 3F13 HIS
SEQADV 3F13 HIS
SEQADV 3F13 HIS
SEQADV 3F13 HIS
SEQADV 3F13 HIS
SEQADV 3F13 MET
SEQADV 3F13 SER
SEQADV 3F13 GLU
SEQADV 3F13 GLY
SEQADV 3F13 HIS
SEQADV 3F13 HIS
SEQADV 3F13 HIS
SEQADV 3F13 HIS
SEQADV 3F13 HIS
SEQADV 3F13 HIS
SEQRES   1 A  16                                                SER
SEQRES   2 A  16                                                PRO
SEQRES   3 A  16                                                TYR
SEQRES   4 A  16                                                ARG
SEQRES   5 A  16                                                GLY
SEQRES   6 A  16                                                ILE
SEQRES   7 A  16                                                ALA
SEQRES   8 A  16                                                ILE
SEQRES   9 A  16                                                PHE
SEQRES  10 A  16                                                LEU
SEQRES  11 A  16                                                ALA
SEQRES  12 A  16                                                GLU
SEQRES  13 A  16
SEQRES   1 B  16                                                SER
SEQRES   2 B  16                                                PRO
SEQRES   3 B  16                                                TYR
SEQRES   4 B  163  ASN LEU PRO GLY GLY LYS ALA ASN ARG GLY GLU LEU ARG
SEQRES   5 B  163  SER GLN ALA LEU ILE ARG GLU ILE ARG GLU GLU THR GLY
SEQRES   6 B  163  LEU ARG ILE ASN SER MET LEU TYR LEU PHE ASP HIS ILE
SEQRES   7 B  163  THR PRO PHE ASN ALA HIS LYS VAL TYR LEU CYS ILE ALA
SEQRES   8 B  163  GLN GLY GLN PRO LYS PRO GLN ASN GLU ILE GLU ARG ILE
SEQRES   9 B  163  ALA LEU VAL SER SER PRO ASP THR ASP MET ASP LEU PHE
SEQRES  10 B  163  VAL GLU GLY ARG ALA ILE LEU ARG ARG TYR ALA ARG LEU
SEQRES  11 B  163  ARG ASN GLU GLU THR ALA LYS GLY GLU ALA LEU ARG ALA
SEQRES  12 B  163  LEU LEU GLY LEU ALA ARG TYR ILE ALA LYS VAL ASP GLU
SEQRES  13 B  163  GLY HIS HIS HIS HIS HIS HIS
FORMUL   3  HOH   *188(H2 O)
HELIX    1   1 LEU A   52  GLY A   66  1                                  15
HELIX    2   2 PHE A  118  ARG A  132  1                                  15
HELIX    3   3 THR A  136  ASP A  156  1                                  21
HELIX    4   4 LEU B   52  GLY B   66  1                                  15
HELIX    5   5 PHE B  118  GLU B  134  1                                  17
HELIX    6   6 THR B  136  HIS B  159  1                                  24
SHEET    1   A 4 GLY A  44  LYS A  46  0
SHEET    2   A 4 ALA A  17  ILE A  23 -1  N  ALA A  20   O  GLY A  45
SHEET    3   A 4 ASN A  83  ILE A  91  1  O  TYR A  88   N  ILE A  23
SHEET    4   A 4 SER A  71  ILE A  79 -1  N  LEU A  75   O  VAL A  87
SHEET    1   B 3 GLU A  25  MET A  26  0
```

A Critical Role of a Cellular Membrane Traffic Protein in Poliovirus RNA Replication

George A. Belov[1], Qian Feng[1*], Krisztina Nikovics[2], Catherine L. Jackson[2], Ellie Ehrenfeld[1*]

Abstract

Introduction

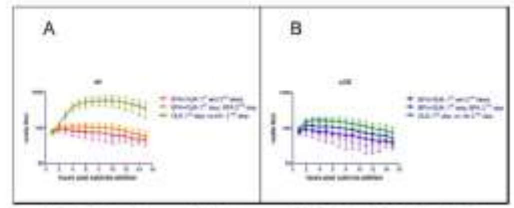


Figure 6. Vesicles formed in the presence of BFA do not support viral RNA replication.

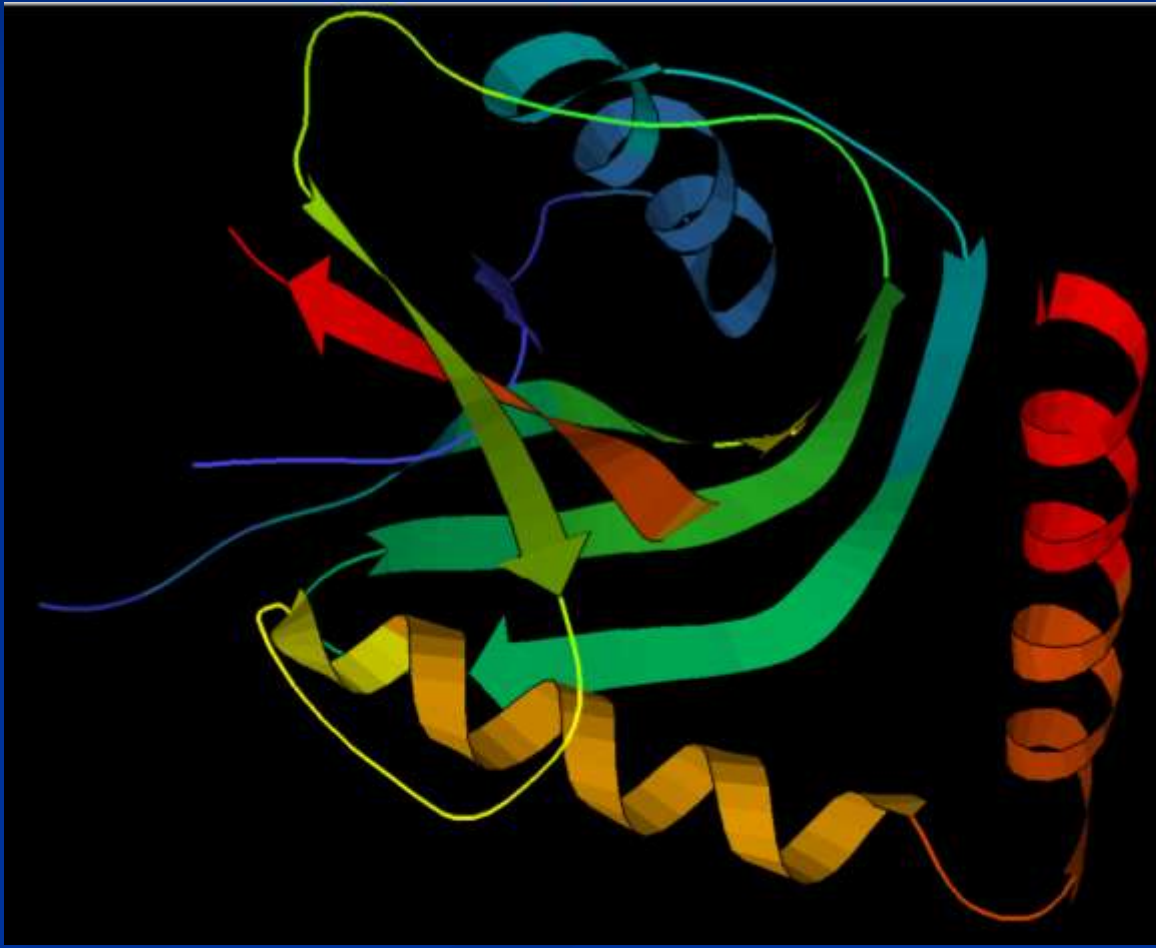
```
ATOM      1  N   ARG A  11      15.440  60.933  40.699  1.00 44.16           N
ATOM      2  CA  ARG A  11      15.197  61.381  39.289  1.00 42.96           C
ATOM      3  C   ARG A  11      15.706  60.349  38.292  1.00 42.40           C
ATOM      4  O   ARG A  11      16.675  59.633  38.561  1.00 42.25           O
ATOM      5  CB  ARG A  11      15.879  62.723  39.019  1.00 43.62           C
ATOM      6  CG  ARG A  11      14.995  63.712  38.289  1.00 45.16           C
ATOM      7  CD  ARG A  11      15.742  64.995  37.971  1.00 44.82           C
ATOM      8  NE  ARG A  11      16.346  65.548  39.166  1.00 45.99           N
ATOM      9  CZ  ARG A  11      17.644  65.495  39.445  1.00 45.13           C
ATOM     10  NH1 ARG A  11      18.474  64.898  38.598  1.00 46.91           N
ATOM     11  NH2 ARG A  11      18.110  66.049  40.571  1.00 45.04           N
ATOM     12  N   LEU A  12      15.051  60.292  37.139  1.00 41.14           N
ATOM     13  CA  LEU A  12      15.427  59.371  36.080  1.00 40.25           C
ATOM     14  C   LEU A  12      16.576  59.931  35.251  1.00 39.81           C
ATOM     15  O   LEU A  12      16.402  60.915  34.519  1.00 40.79           O
ATOM     16  CB  LEU A  12      14.219  59.105  35.179  1.00 40.07           C
ATOM     17  CG  LEU A  12      14.539  58.258  33.941  1.00 38.13           C
ATOM     18  CD1 LEU A  12      14.795  56.773  34.381  1.00 37.07           C
ATOM     19  CD2 LEU A  12      13.419  58.363  32.929  1.00 39.09           C
ATOM     20  N   PRO A  13      17.764  59.310  35.344  1.00 39.27           N
ATOM     21  CA  PRO A  13      18.859  59.757  34.498  1.00 38.54           C
ATOM     22  C   PRO A  13      18.523  59.598  33.005  1.00 37.53           C
ATOM                                                                         O
ATOM                                                                         C
ATOM                                                                         C
ATOM                                                                         C
ATOM                                                                         N
ATOM                                                                         C
ATOM                                                                         C
ATOM                                                                         O
ATOM                                                                         C
ATOM                                                                         O
ATOM                                                                         N
ATOM                                                                         C
ATOM                                                                         C
ATOM                                                                         O
ATOM                                                                         C
ATOM                                                                         C
ATOM                                                                         O
ATOM                                                                         O
ATOM                                                                         N
ATOM                                                                         C
ATOM                                                                         C
ATOM                                                                         O
ATOM                                                                         C
ATOM                                                                         C
ATOM                                                                         C
ATOM                                                                         C
ATOM                                                                         N
ATOM                                                                         C
ATOM                                                                         C
ATOM                                                                         O
ATOM                                                                         C
ATOM                                                                         N
ATOM                                                                         C
ATOM                                                                         C
ATOM                                                                         O
ATOM                                                                         C
ATOM     59  CG  ARG A  18                                                   C
ATOM     60  CD  ARG A  18      30.553  56.236  26.207  1.00 31.69           C
ATOM     61  NE  ARG A  18      31.769  55.538  25.853  1.00 32.21           N
ATOM     62  CZ  ARG A  18      33.004  56.038  25.819  1.00 31.93           C
ATOM     63  NH1 ARG A  18      33.262  57.321  26.076  1.00 31.65           N
```

CRYSTAL STRUCTURE OF PUTATIVE NUDIX HYDROLASE FAMILY MEMBER FROM CHROMOBACTERIUM VIOLACEUM

J.B.BONANNO,J.FREEMAN,
K.T.BAIN, J.DO,R.ROMERO,
S.WASSERMAN, J.M.SAUDER,
S.K.BURLEY,S.C.ALMO,

NEW YORK SGX
RESEARCH CENTER
FOR STRUCTURAL
GENOMICS (NYSGXRC)

27-OCT-08



Source: RCSB Protein Data Bank; www.rcsb.org

Redefining “Computer”

- “All the calculations that would ever be needed in this country could be done on three digital computers."

  -- Douglas Hartree, Cambridge, 1954

- "There is no reason for an individual to have a computer in their home."

  -- Ken Olsen, DEC, 1977

- “For the full year [2007], IDC said 269 million PCs were shipped worldwide”

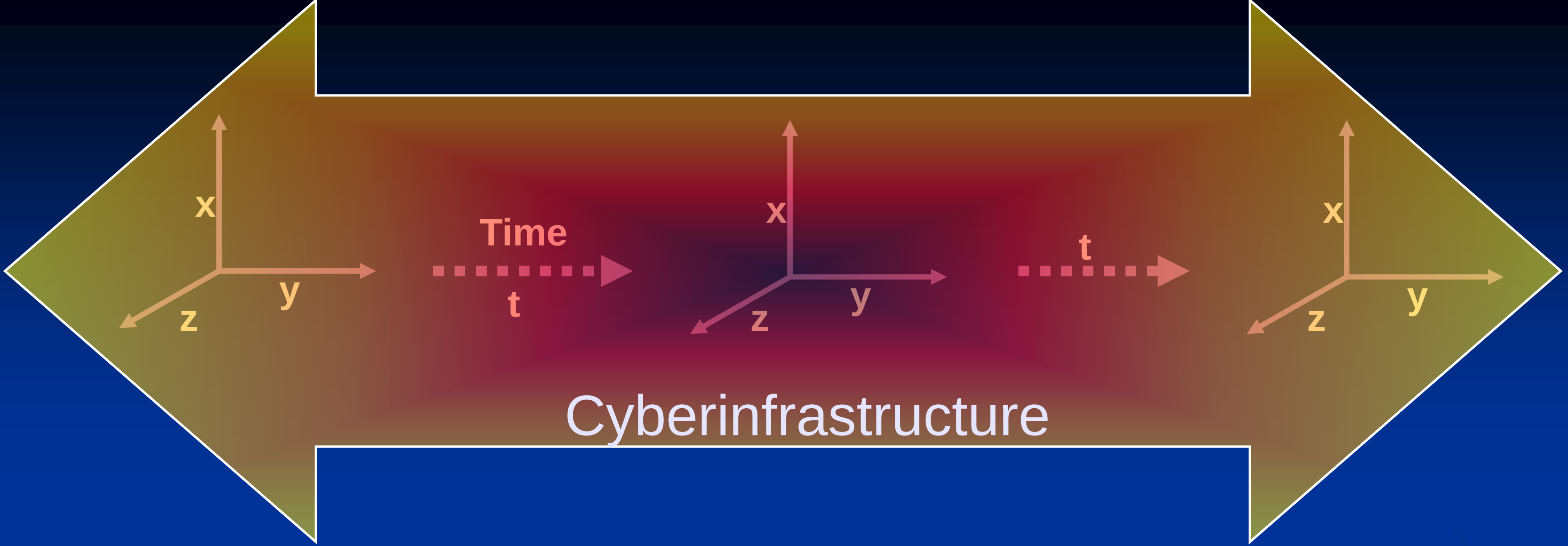
  -- International Herald Tribune, January, 2008.

- "In a sense, there are only five computers on earth."

  -- Yahoo Research Chief Prabhakar Raghavan, December, 2007.

- “…some researchers at IBM believe that five computers may be four too many.”

  -- Nick Carr, The Guardian, February 21, 2008.

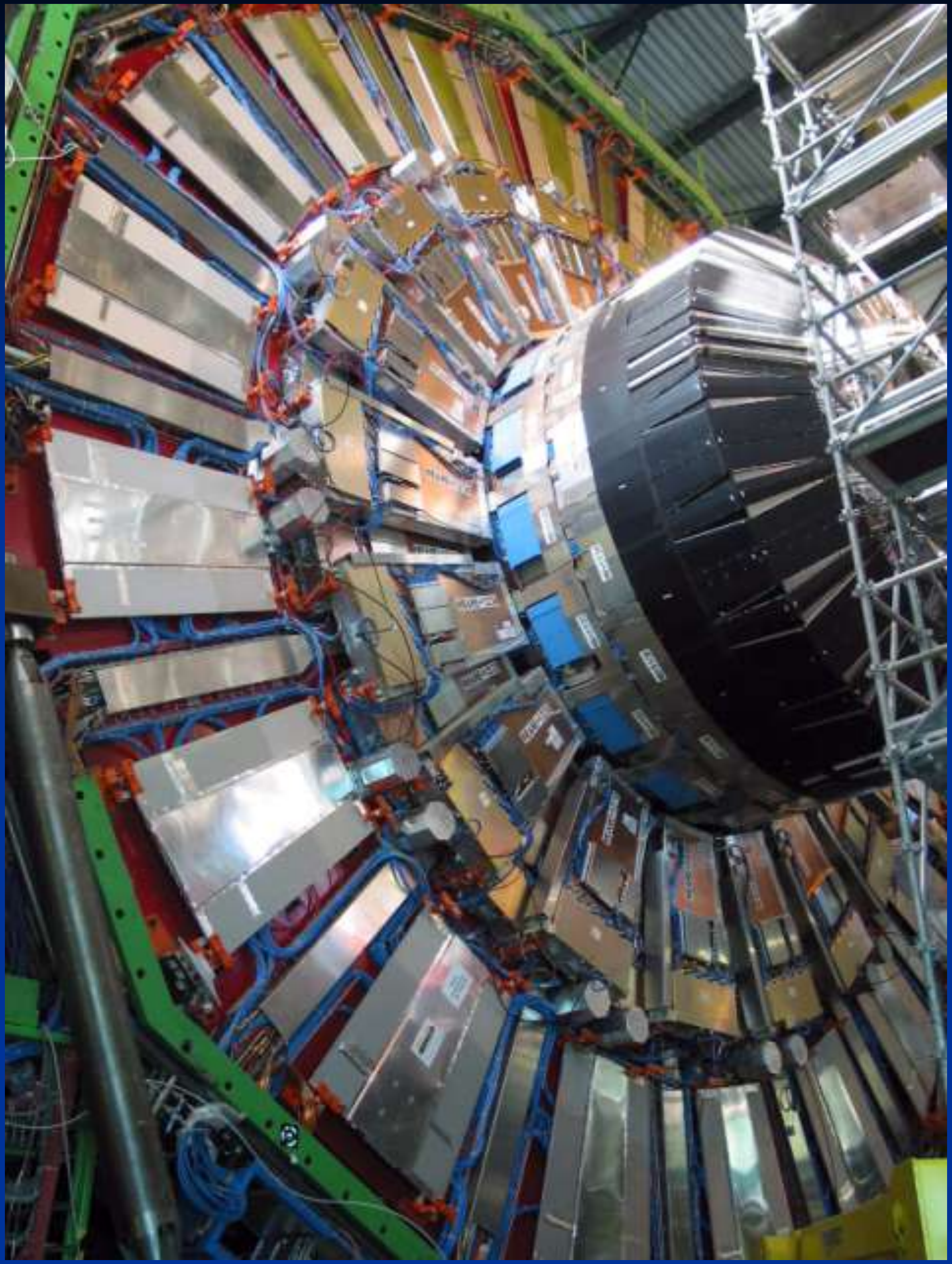


Information for innovation and discovery


LSST
Large Synoptic Survey Telescope

“Sometime in the 2010s, if all goes well, the Large Synoptic Survey Telescope (LSST) will start to bring a vision of the heavens to Earth. Suspended between its vast mirrors will be a three billion-pixel sensor array, which on a clear winter night will produce 30 terabytes of data. In less than a week this remarkable telescope will map the whole night sky …. And then the next week it will do the same again … building up a database of billions of objects and millions of billions of bytes.”

Nature 440:383



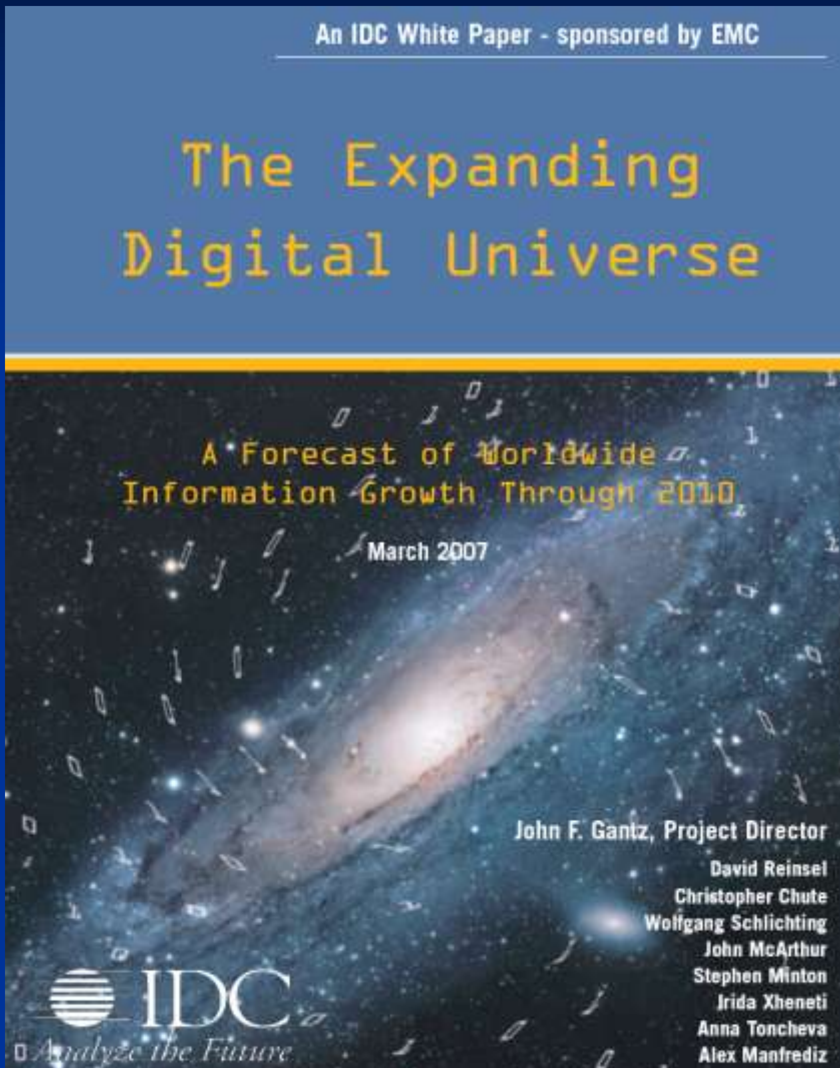
Large Hadron Collidor

Physicists will use the LHC to recreate the conditions just after the Big Bang, by colliding two beams [of hadrons] head-on at very high energy.

Source: public.web.cern.ch/Public/en/LHC/LHC-en.html

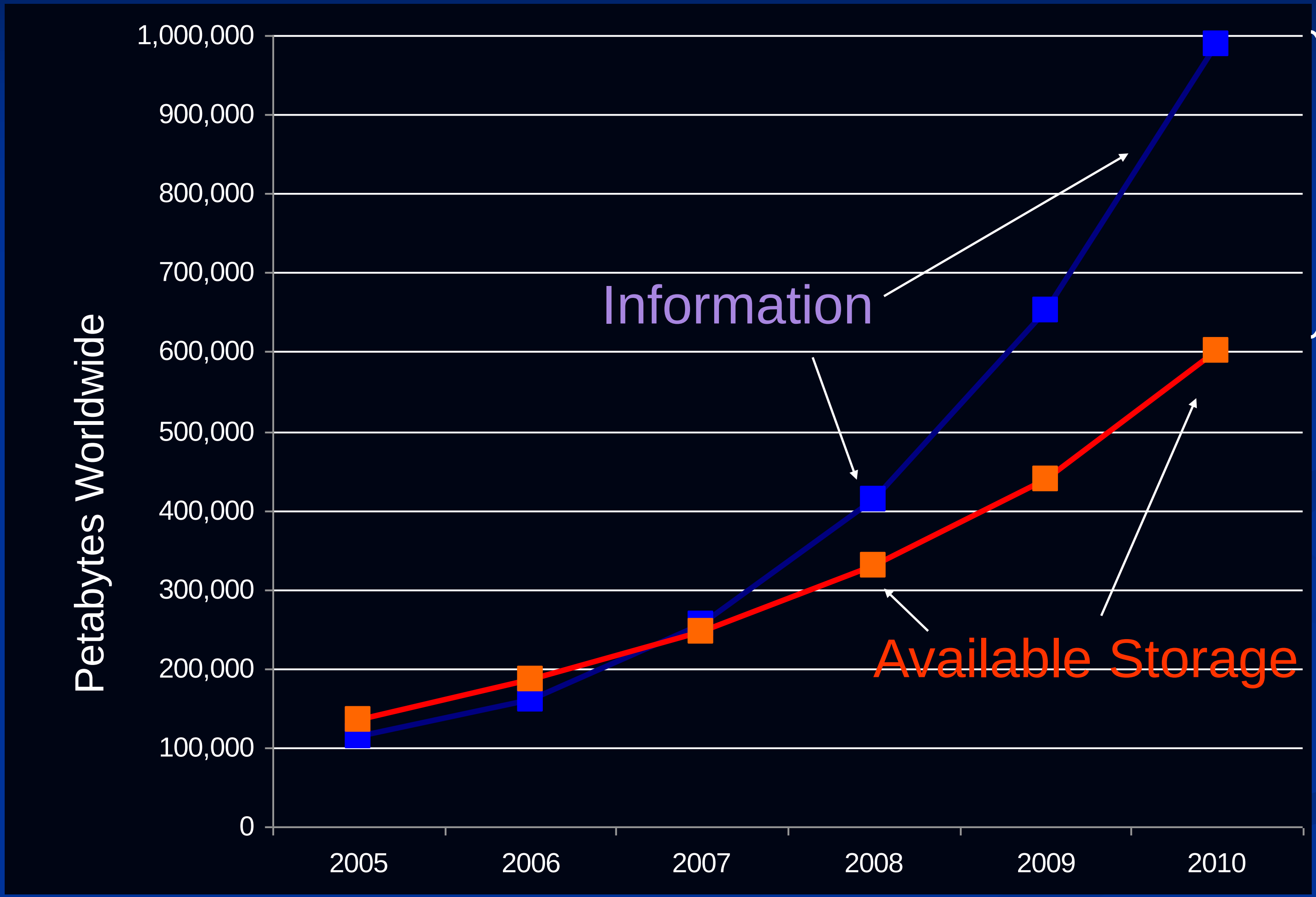
When LHC begins operations, it will produce roughly 15 Petabytes of data annually, which thousands of scientists around the world will access and analyse … The mission of the LHC Computing Project (LCG) is to build and maintain a data storage and analysis infrastructure for the entire high energy physics community that will use the LHC.

Source: public.web.cern.ch/Public/en/LHC/LHC-en.html



"In 2006, the amount of digital information created, captured, and replicated was 1,288 x 10^{18} bits (or 161 exabytes) … This is about 3 million times the information in all the books ever written"

Information And Storage



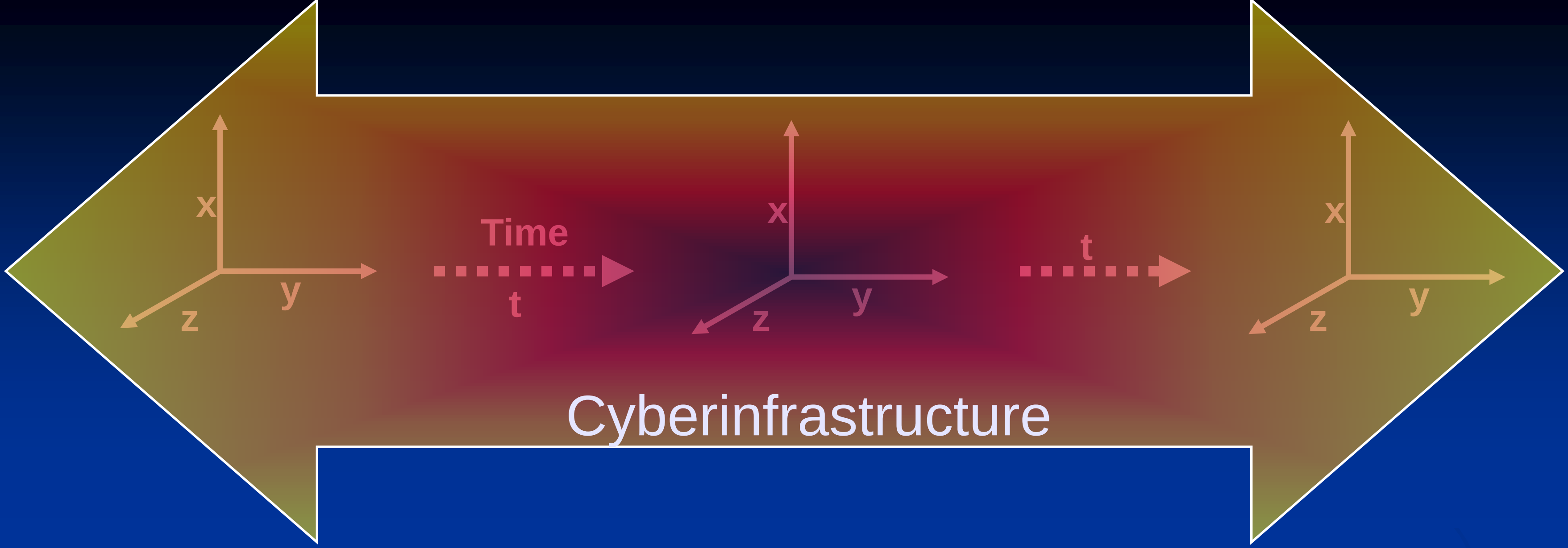
Transient information or unfilled demand for storage

Source: John Gantz, IDC Corporation
The Expanding Digital Universe

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

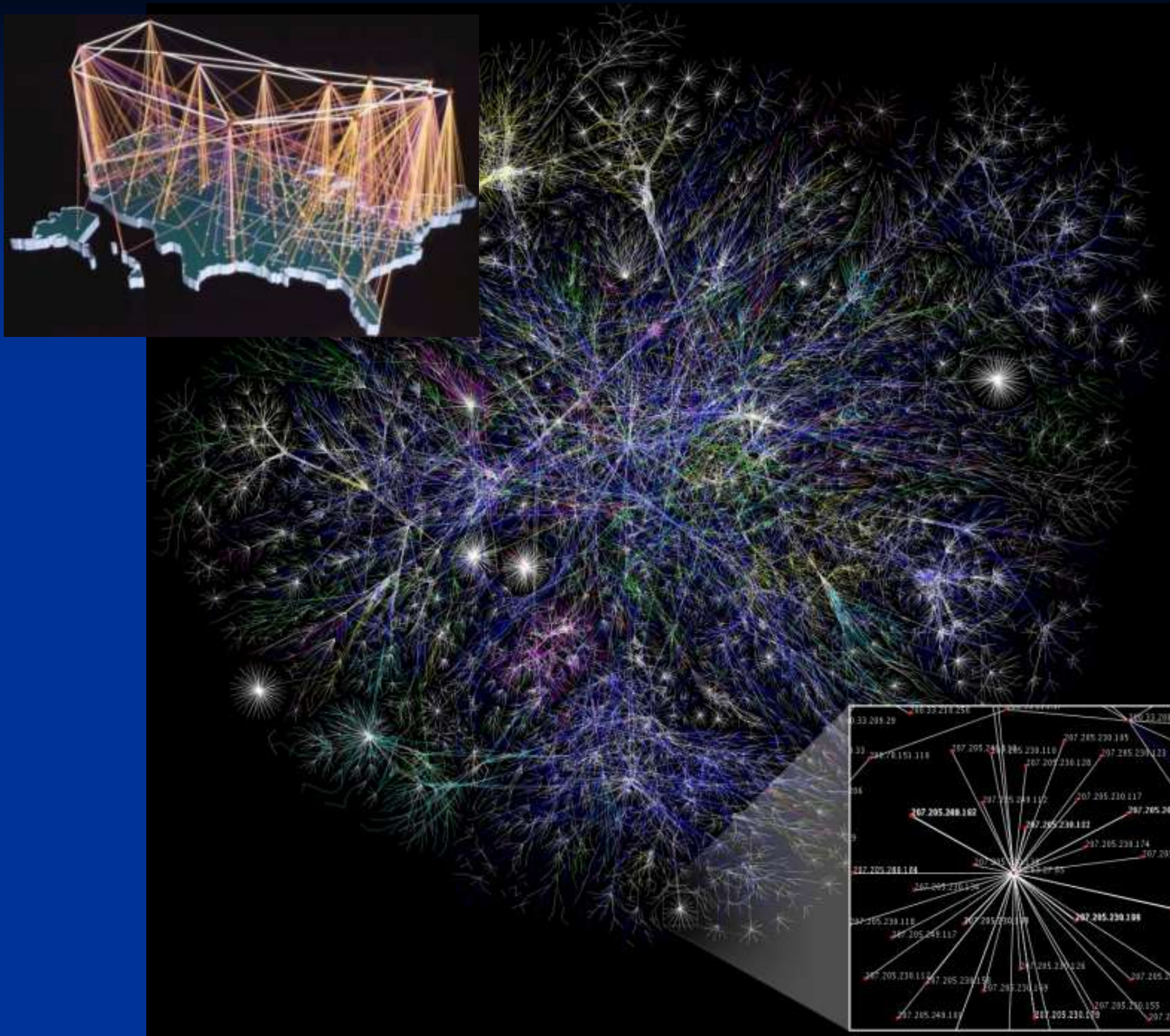
Google's founding philosophy is that we don't know why this page is better than that one: If the statistics … say it is, that's good enough. No semantic or causal analysis is required. That's why Google can translate languages without actually "knowing" them (given equal corpus data, Google can translate Klingon into Farsi as easily as it can translate French into German).

- Chris Anderson   Wired Magazine 06.23.08

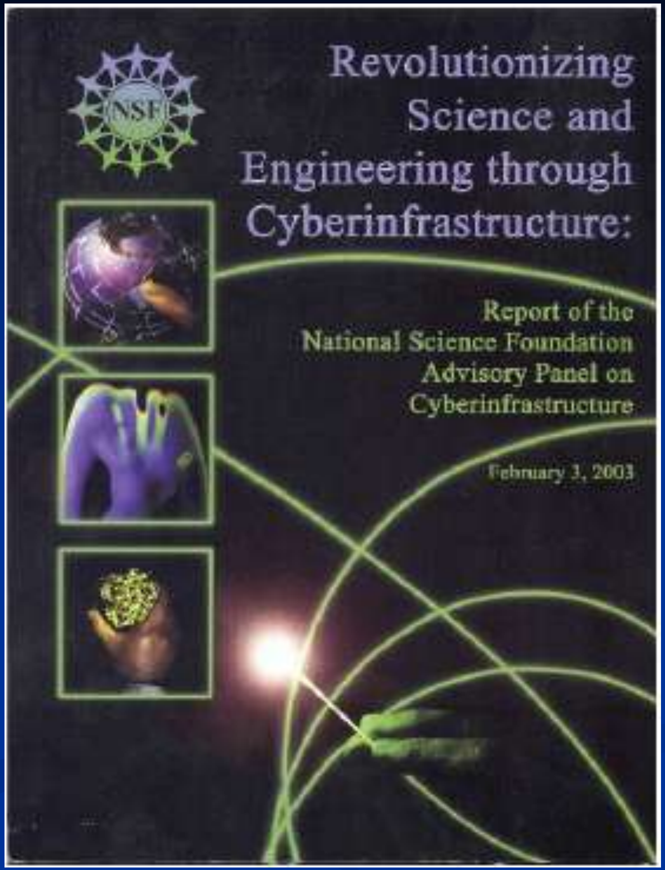


Connectivity
for access and
interaction

The Department of Defense’s ARPANET project, launched in 1966 to explore methods for “resource sharing among computers”, initially connected 4 nodes. Today’s Internet links more than 1.4 billion users over more than 200,000 networks worldwide; with 14 new users added every second.



"New types of scientific organizations [that] *serve individuals, teams and organizations in ways that revolutionize …*

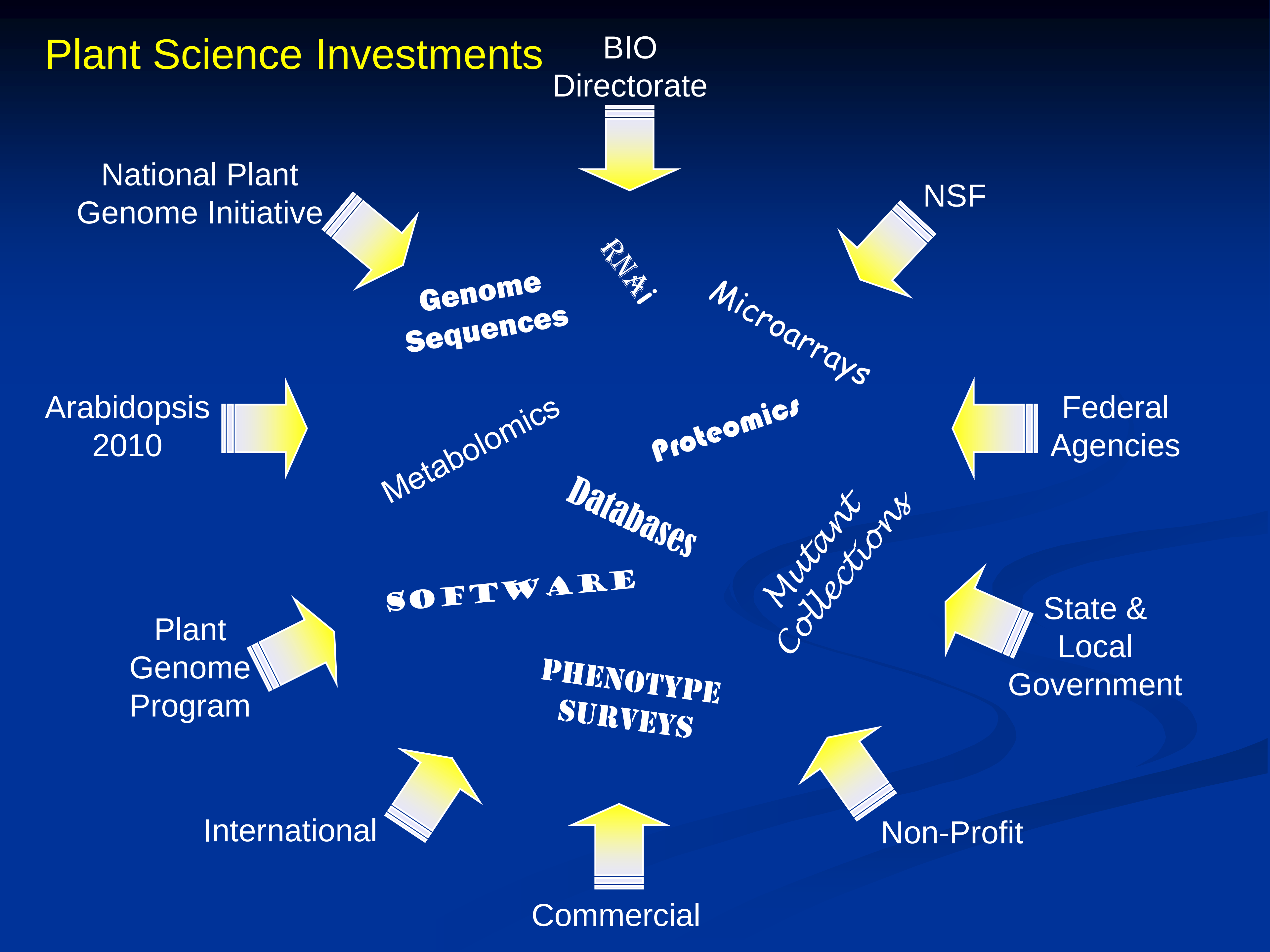


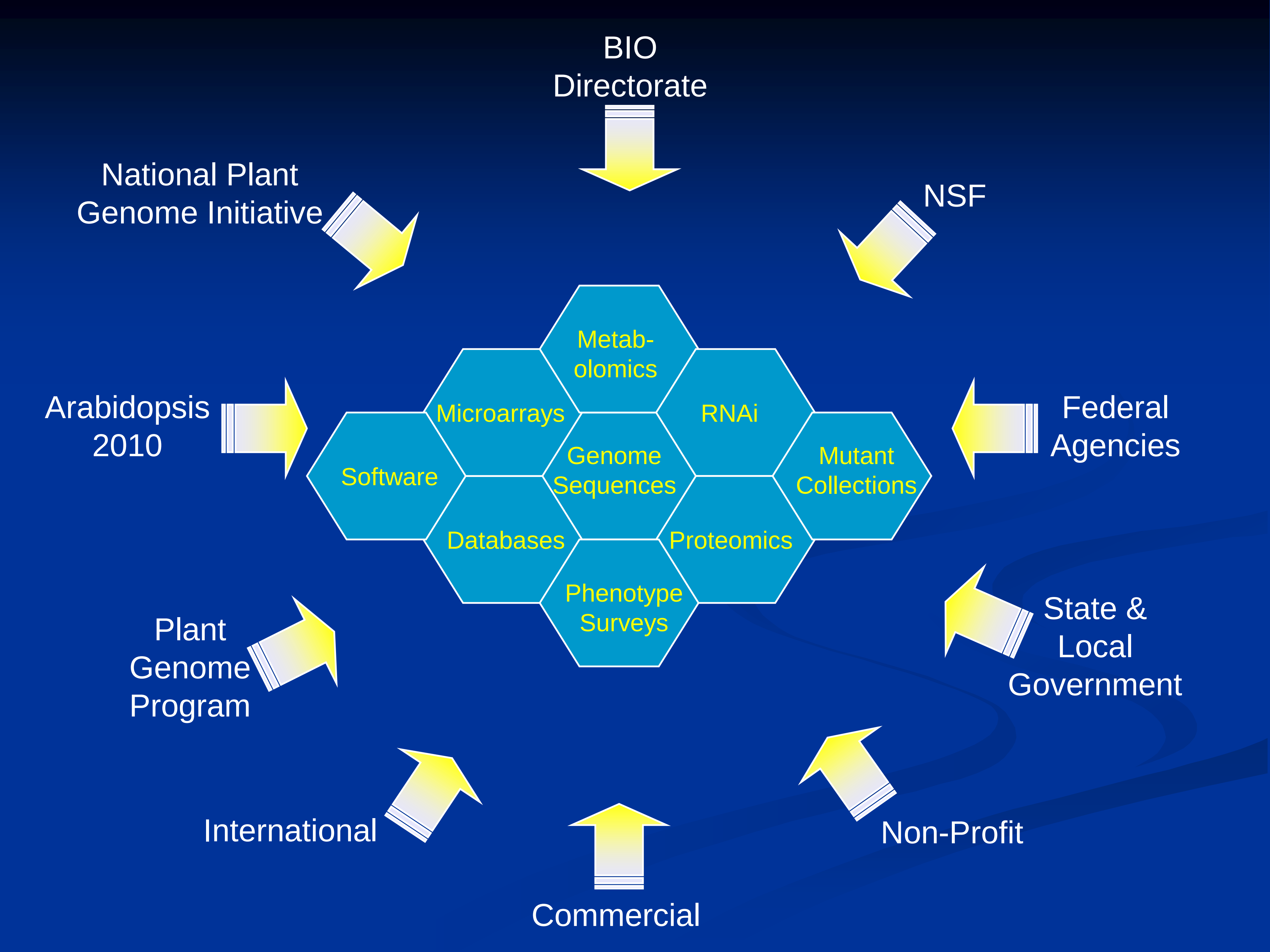
… what they can do, how they do it, and who participates."

iPlant™ Collaborative

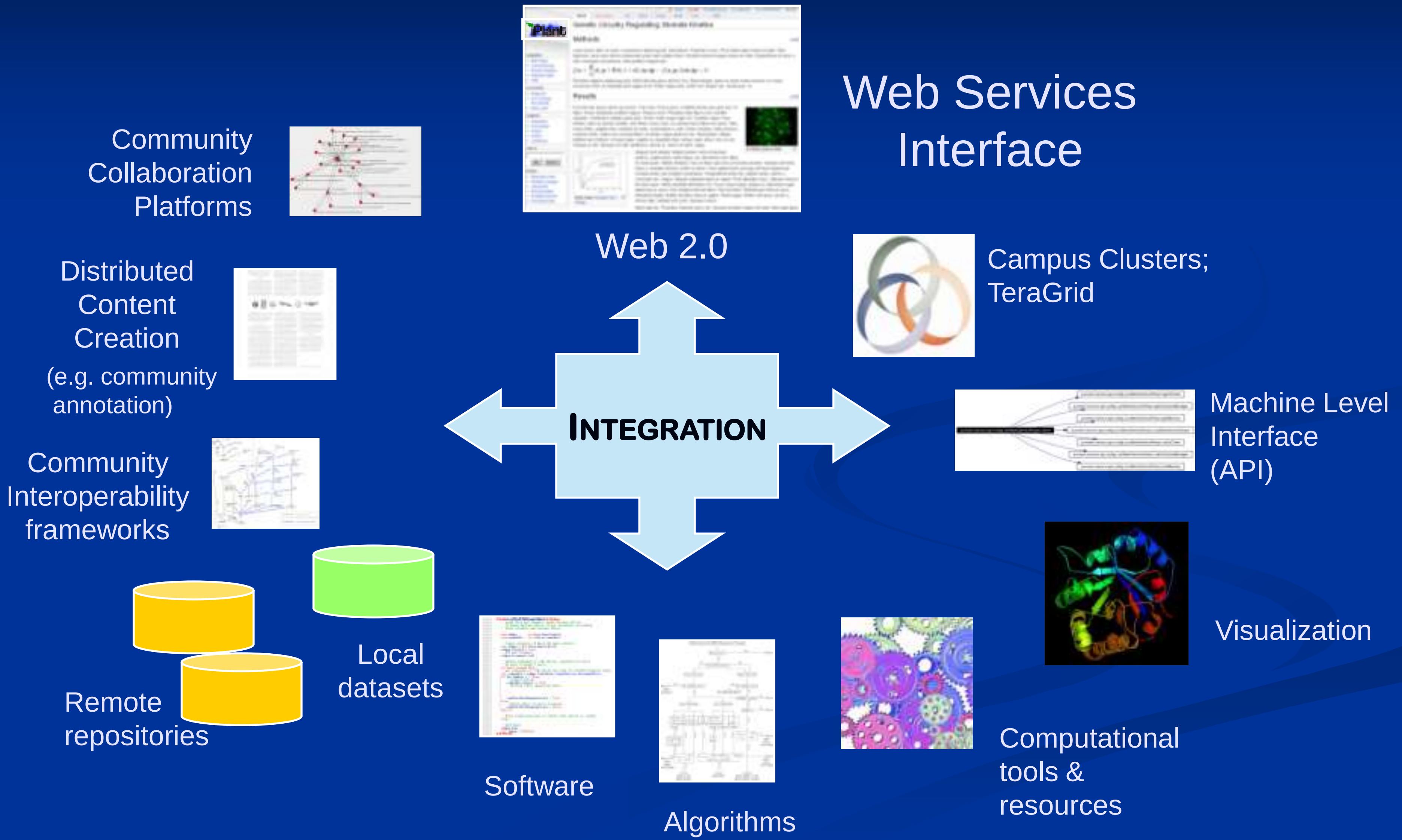
Empowering a new plant biology

- Richard Jorgensen, Collaborative Director, Plant Sciences/BIO5, University of Arizona
- Steve Goff, Director of Community Interactions, University of Arizona
- Rick Blevins, Director of Cyberinfrastructure Development, University of Arizona
- Martha Narro, Director of Education, Outreach and Training, University of Arizona
- Greg Andrews, Team Leader, Computing Infrastructure Team, Computer Science, University of Arizona
- Sudha Ram, Co-Team Leader, Integrated Solutions Team, Management Information Systems, University of Arizona
- Lincoln Stein, Co-Team Leader, Integrated Solutions Team, Bioinformatics, Cold Spring Harbor Laboratory
- Steve Rounsley, Team Leader, Synthesis Activities Team, BIO5, University of Arizona
- Vicki Chandler, Team Leader, Education, Outreach and Training Team, Director of BIO5, University of Arizona
- Susan Brown, Team Leader, Social Science Team, Management Information Systems, University of Arizona
- Kobus Barnard, Computer Science, University of Arizona
- Travis Huxman, Ecology and Evolutionary Biology, University of Arizona
- Dan Stanzione, Director, Fulton High Performance Computing Center, Arizona State University
- Ann Stapleton, Biology, University of North Carolina-Wilmington
- Matt Vaughn, Bioinformatics, Cold Spring Harbor Laboratory
- Doreen Ware, Bioinformatics, USDA/Cold Spring Harbor Laboratory


Plant Science Investments
BIO Directorate
National Plant Genome Initiative
NSF
Arabidopsis 2010
Federal Agencies
Plant Genome Program
State & Local Government
International
Commercial
Non-Profit
Genome Sequences
RNAi
Microarrays
Metabolomics
Proteomics
Databases
Mutant Collections
SOFTWARE
PHENOTYPE SURVEYS

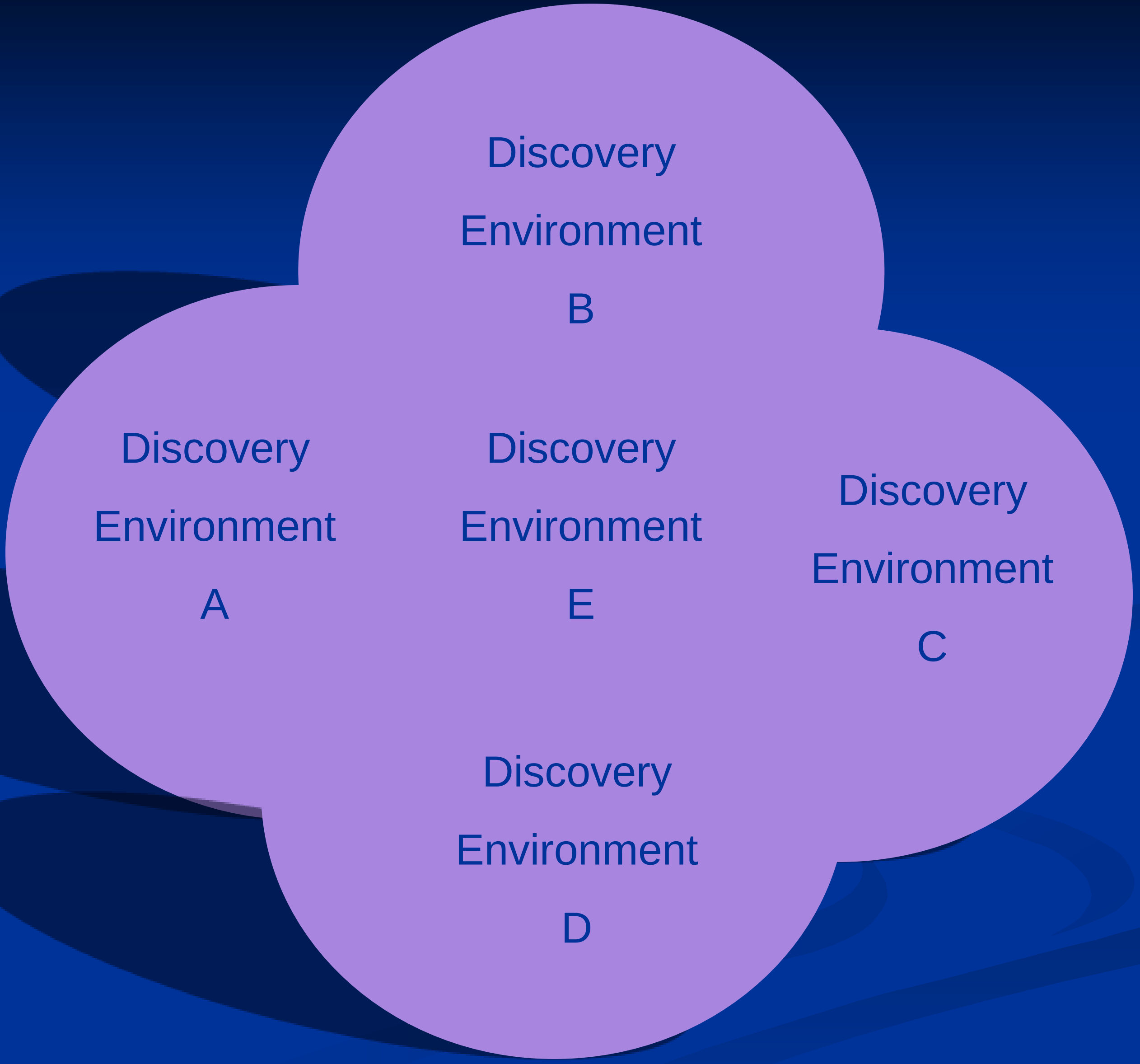

BIO Directorate
National Plant Genome Initiative
NSF
Arabidopsis 2010
Federal Agencies
Plant Genome Program
State & Local Government
International
Commercial
Non-Profit
Metab-olomics
Microarrays
RNAi
Software
Genome Sequences
Mutant Collections
Databases
Proteomics
Phenotype Surveys

A Discovery Environment



Discovery Landscape

- Distributed
- Integrated
- Open source
- Evolvable
- Extensible



The Future: Next Generation Networking and Information Technology

- **Emergent Cognition**
- **Beyond Virtual**
- **Pathways to Trust and Confidence**

Emergent Cognition

Harness new levels of intelligence, intuition and perception that emerge as greater than the sum of the parts in the union of cyber, human, and social capabilities

Example: Enable the combined use of logic, analysis, and modeling; expert intuition and perception; and market judgment and crowd wisdom in understanding the properties and behavior of a global economy.

Beyond Virtual

Extend the envelope of capabilities by making the virtual and the physical one.

Example: Cognitive computing – Provide the ad hoc capability to access the full sensing, information, and computational capabilities of the every day objects around you to respond to an emergency or to an ordinary need.

Pathways to Trust and Confidence

Ensure the value to society of cyber systems is not limited by a lack of trust and confidence.

Example: Provide capabilities for identity management, information security, privacy protection, and tiered access to enable an electronic health records system that can be confidently used for individual patient care, public health, and biomedical research.

eLibrary - Single point of access connecting resources worldwide:

- Digital and physical repositories
- Computational capabilities
- Digital environments for access and interaction by people, machines, and organizations
- Software tools and resources

Contact information:

greer@nitrd.gov