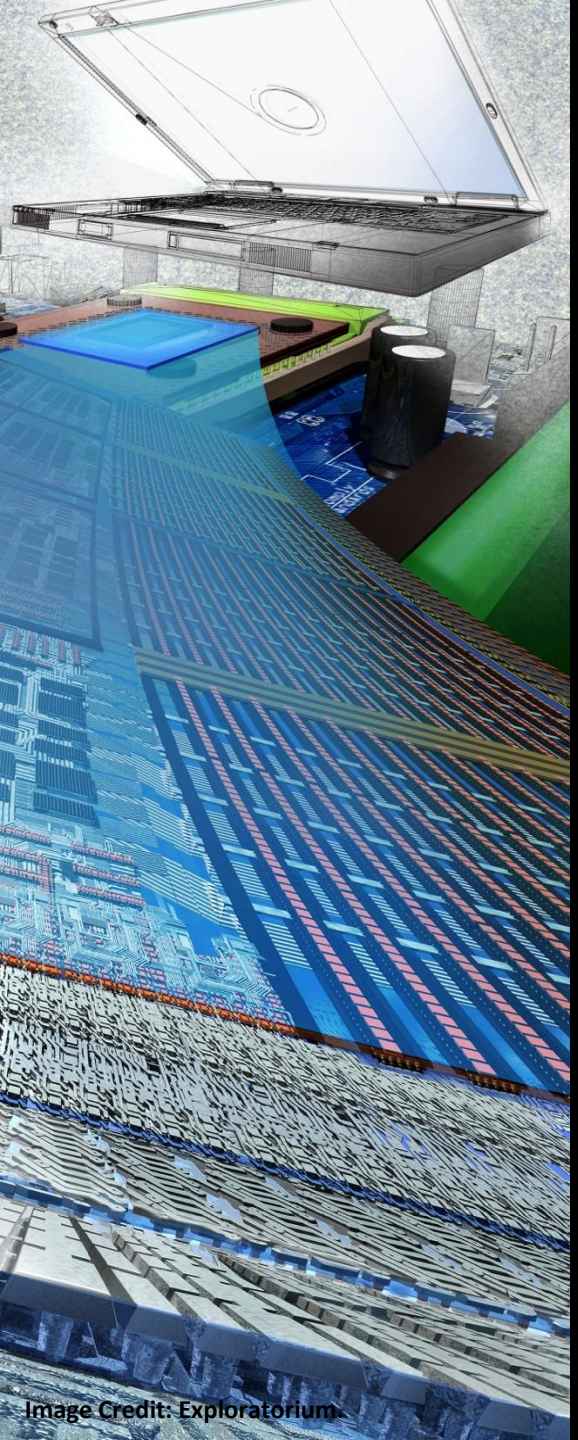


Big Data R&D Initiative

Howard Wactlar
CISE Directorate
National Science Foundation

NIST Big Data Meeting
June, 2012

The Landscape: Smart Sensing, Reasoning and Decision

Environment Sensing





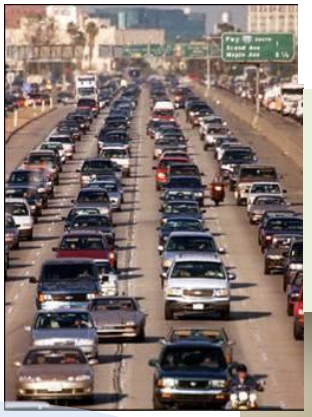


Percepts (sensors)

Agent (Reasoning)

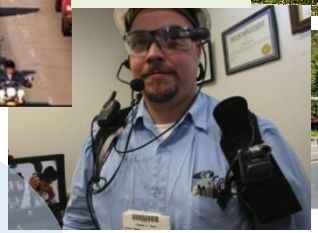
Actions (controllers)

Emergency Response



Situation Awareness: Humans as sensors feed multi-modal data streams

Credit: Photo by US Geological Survey




Pervasive Computing

Social Informatics

People-Centric Sensing

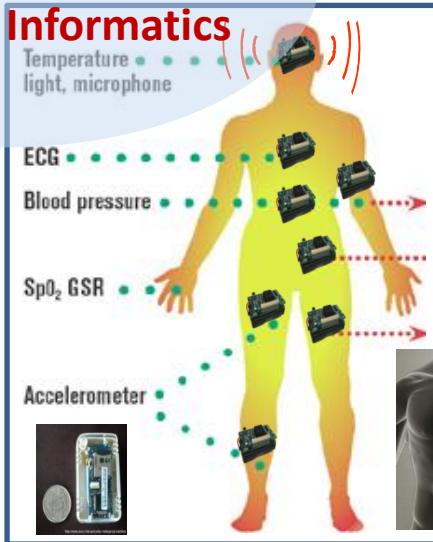



Public Sensing

Social Sensing

Personal Sensing

Smart Health Care



Temperature, light, microphone

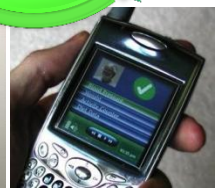
ECG

Blood pressure

SpO_2 GSR

Accelerometer

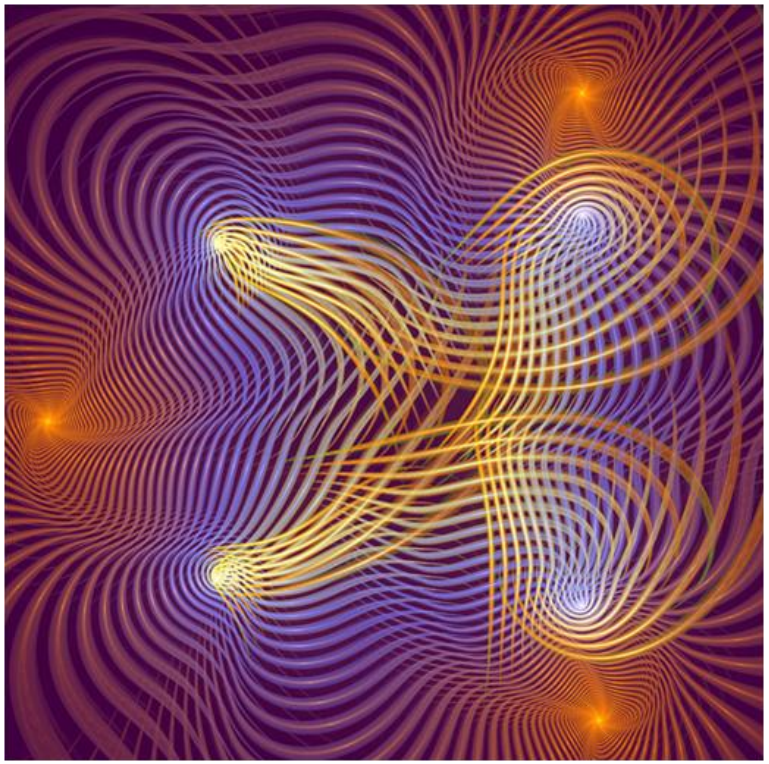
Evaluate, Sense, Intervene, Identify, Assess

Source: Sajal Das, Keith Marzullo

Paradigm Shift: from Hypothesis-driven to Data-driven Discovery

- **Data-driven discovery is** revolutionizing scientific exploration and engineering innovations
- **Automatic extraction of new knowledge** about the physical, biological and cyber world continues to accelerate
- Multi-cores, concurrent and parallel algorithms, virtualization and advanced server architectures will enable **data mining and machine learning**, and **discovery and visualization of Big Data**



Examples of Research Challenges

- More data is being collected than we can store
  - Analyze the data as it becomes available
  - Decide what to archive and what to discard
- Many data sets are too large to download
  - Analyze the data wherever it resides
- Many data sets are too poorly organized to be usable
  - Better organize and retrieve data
- Many data sets are heterogeneous in type, structure, semantics, organization, granularity, accessibility …
  - Integrate and customize access to federate data
- Utility of data limited by our ability to interpret and use it
  - Extract and visualize actionable knowledge
  - Evaluate results
- Large and linked datasets may be exploited to identify individuals
  - Design management and analysis with built-in privacy preserving characteristics

Administration's Big Data Research and Development Initiative

- Big Data Senior Steering Group – chartered in spring 2011 under the Networking and Information Technology R&D (NITRD) Program
  - Members from NIH, NSF, DARPA, DOD OSD, DHS, DOE-Science, HHS, NARA, NASA, NIST, NOAA, NSA, and USGS
  - Co-chaired by NIH and NSF

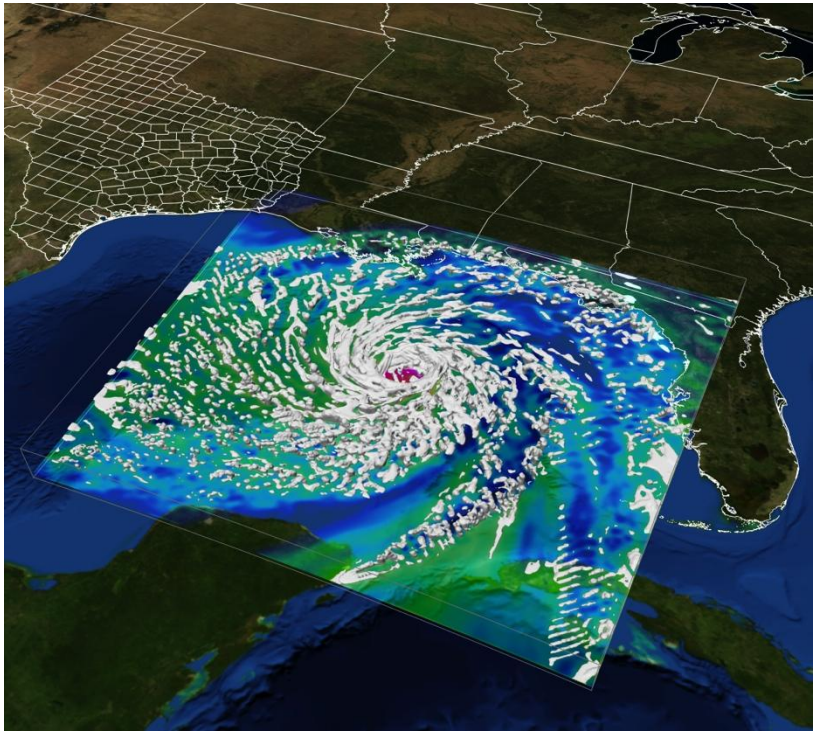
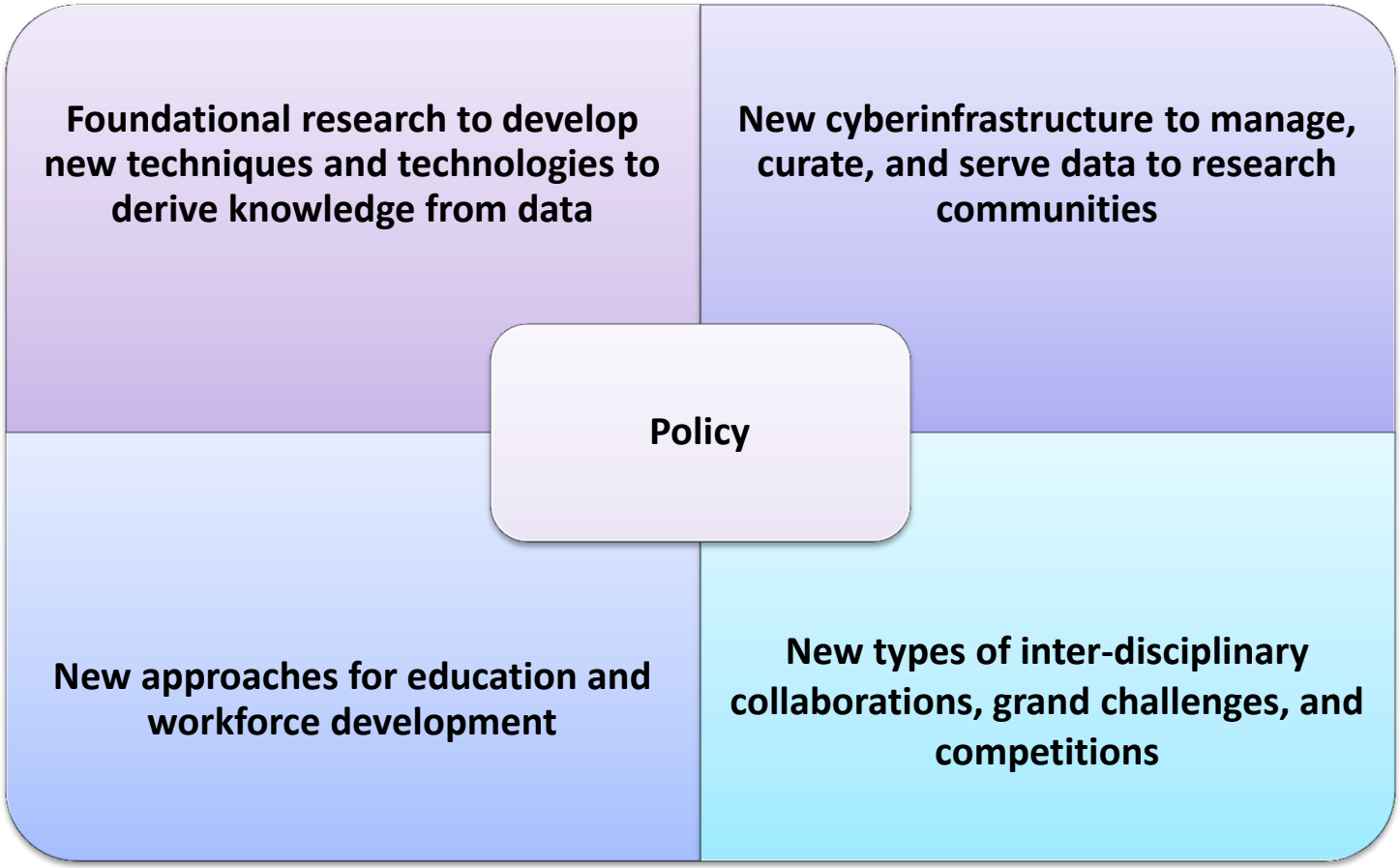


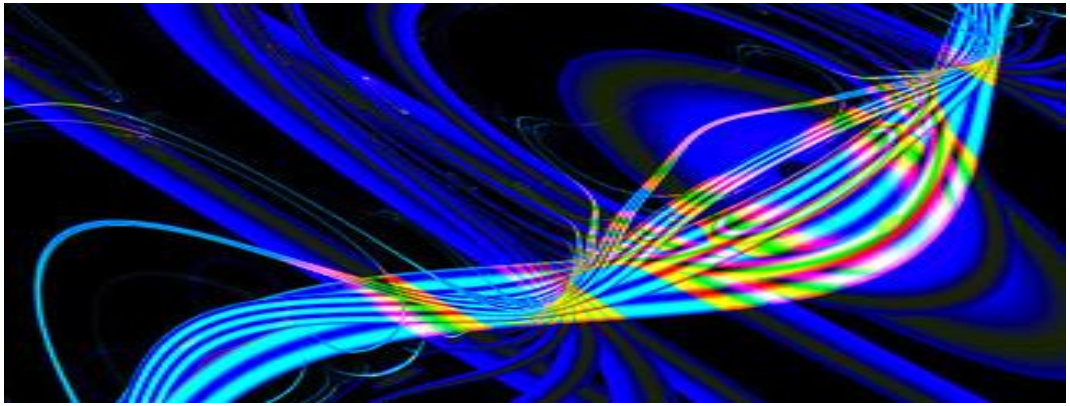
Image Credit: *Fuqing Zhang and Yonghui Weng, Pennsylvania State University; Frank Marks, NOAA; Gregory P. Johnson, Romy Schneider, John Cazes, Karl Schulz, Bill Barth, The University of Texas at Austin*

NSF Strategy to Address Big Data



A Complex Policy Setting

- Researchers want data.
- Public policy requires access to data.
- Public policy also requires protection of privacy and intellectual property and other sensitive information.
- Much more to be done: Policy on data management and data access



Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIG DATA)

Foundational research to extract knowledge from data

Foundational research to advance the core techniques and technologies for managing, analyzing, visualizing, and extracting useful information from large, diverse, distributed and heterogeneous data sets.

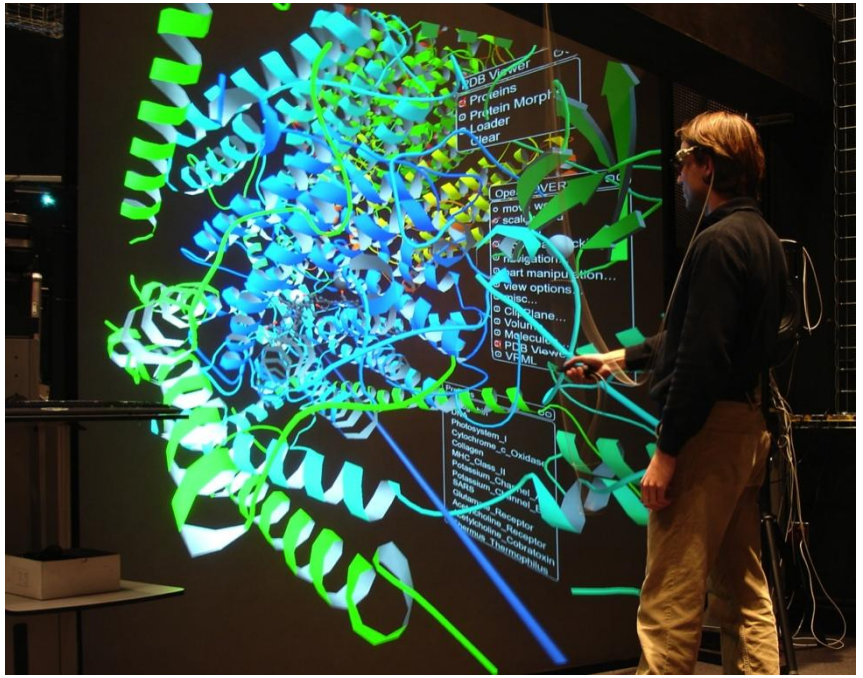


Image Credit: *Jurgen Schulze, Calit2, UC-San Diego*

Cross-Directorate Program: NSF Wide

Multi-agency Commitment: NSF and NIH

BIG DATA Research Thrusts

Collection, Storage, and Management of "Big Data"

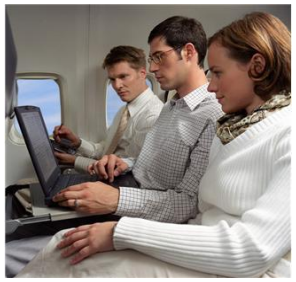
- Data representation, storage, and retrieval
- New parallel data architectures, including clouds
- Data management policies, including privacy and access
- Communication and storage devices with extreme capacities
- Sustainable economic models for access and preservation

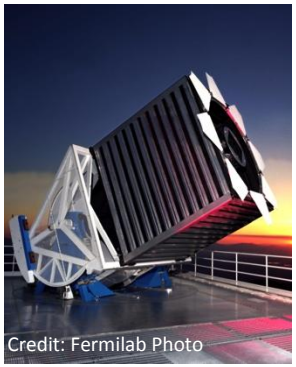
Data Analytics

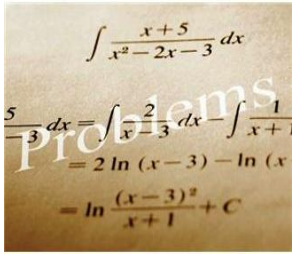
- Computational, mathematical, statistical, and algorithmic techniques for modeling high dimensional data
- Learning, inference, prediction, and knowledge discovery for large volumes of dynamic data sets
- Data mining to enable automated hypothesis generation, event correlation, and anomaly detection
- Information infusion of multiple data sources

Research in Data Sharing and Collaboration

- Tools for distant data sharing, real time visualization, and software reuse of complex data sets
- Cross disciplinary model, information and knowledge sharing
- Remote operation and real time access to distant data sources and instruments

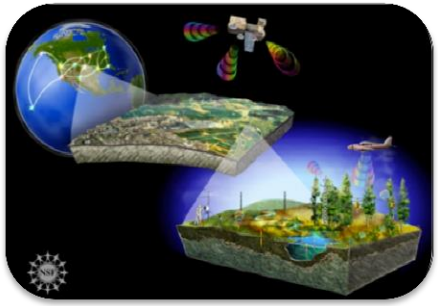




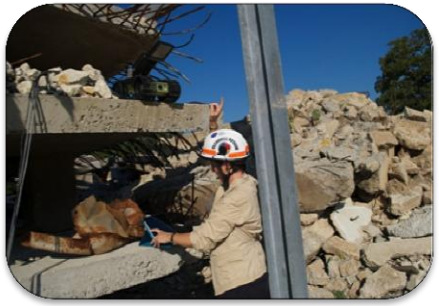


Big Data to Address National Priorities

Health & Wellbeing



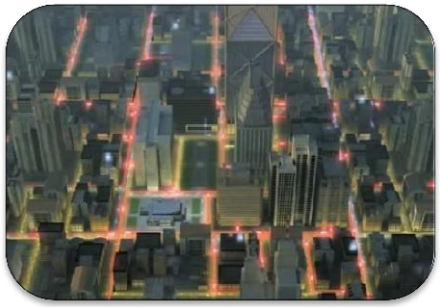
Environment & Sustainability



Emergency Response & Disaster Resiliency

Manufacturing, Robotics, & Smart Systems

Secure Cyberspace



Transportation & Energy

Education and Workforce Development

Smart Health & Wellbeing

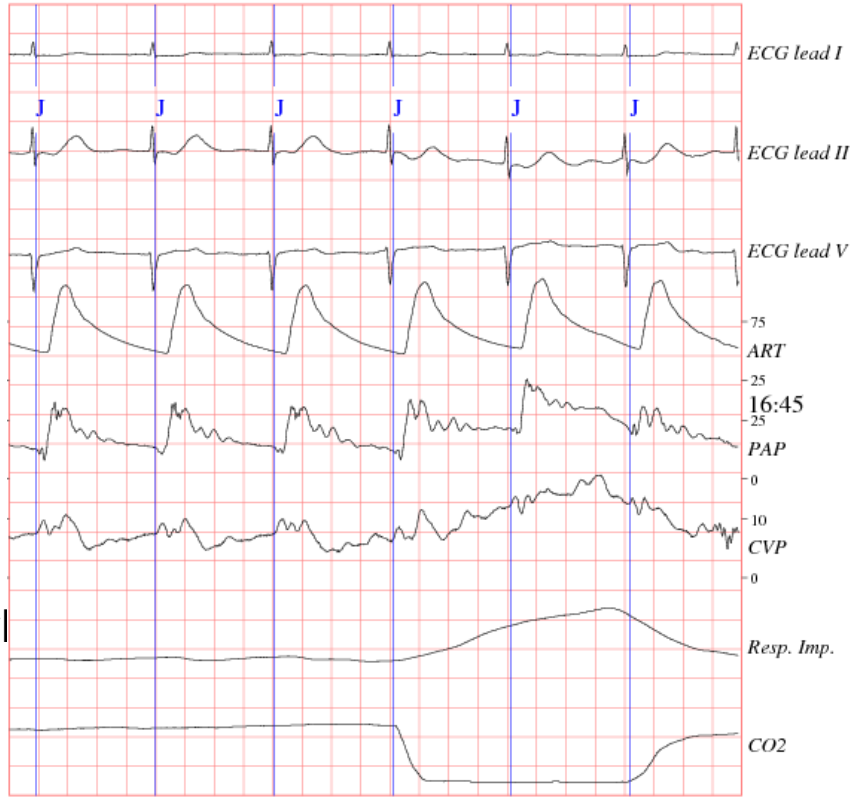
Reasoning under uncertainty

The ability to acquire, aggregate and mine clinical, scientific, behavioral data will create an unprecedented amount of high quality data from individuals and population

Enabling evidence-based medicine, early diagnoses, personalized assessments and care

Era of “Big Data” in Healthcare

- **Large volumes of data currently collected**
  - **EHRs and PHRs**
    - **Multi-scale and multi-source**
  - **During hospitalizations**
    - **For safety and diagnosis**
  - **On an out-patient basis**
    - **Typically event monitors**
  - **Via ubiquitous mobile sensors**
    - **Behavior, physiology, environment**
  - **As part of clinical studies**
    - **To evaluate safety and efficacy**
  - **From growing body of scientific knowl**
    - **In biomedical research literature**



- **Gigabits/patient/day**
  - **High sampling rates**
  - **Multiple signals**
- **Accumulating data is getting easier, but using data is hard**

Secure & Trustworthy Cyberspace

Securing our Nation's Data & Com

New interdisciplinary program that aims to support fundamental scientific advances and technologies to protect cyber-systems from malicious behavior, while preserving privacy and promoting usability.

Cyberlearning: Transforming Education

Integrate technology with knowledge about how people learn

Goals:

- Understand how people learn in technology rich environments
- Design and study ways in which innovative technologies and tools can promote learning and support assessment
- Prototype new technologies and integrate them into learning environments

Big Opportunities for the Future

- Our investments in **research and education** have returned exceptional dividends to our nation.

- Scientific discovery and technological innovation are at the core of our response to **national and societal challenges** – from environment, energy, transportation, sustainability and healthcare, to cyber security and national defense.

- Many of tomorrow's breakthroughs will occur at the **intersections of diverse disciplines**.

Thanks!

hwactlar@nsf.gov