NITRD: National Big Data Strategic Plan

Summary of Request for Information Responses
Introduction: Demographics

Summary of Responses

- Next generation Capabilities
- Data to Knowledge to Action
- Access to Big Data Resources
- Education and Training
- New Partnerships
- Data Sustainability
- Gateways

Additions and Comments

- Gaps
- Game Changers
- Security and Privacy
National Strategic Plan RFI

- We received 38 responses
- From 27 stakeholders including:
  - Industry
  - Academia
  - National Labs
  - Nonprofits
  - Community Leaders
  - Consortium
Response Locations
Next Generation Capabilities

Vision 1: Create next generation capabilities by leveraging emerging Big Data foundations, technologies, processes, and policies. (62%)

- **Data Efficiency***
  - “Algorithm Acceleration with state-of-the-art hardware”
- **Infrastructure/Hardware***
  - “Infrastructure R&D should address delivery of information through varied, multiple, and concurrent channels and mobile devices”
- **Testbeds***
  - “Create cloud-based test beds for R&D with sample datasets, community supported algorithm libraries, with cross-agency access and support.”
- **Data co-design***
  - “…design process where scientific problem requirements influence architecture design and technology, and constraints inform formulation and design of algorithms and software.”

*Multiple responses*
Data to Knowledge to Action

Vision 2: In addition to supporting the R&D necessary to create knowledge from data, emphasize support of R&D to understand trustworthiness of data and resulting knowledge, and to make better decisions and breakthrough discoveries and take confident action based on them. (46%)

• Time to Knowledge
  • “...all processing could be done “inline,” with the data stream rather than as a post processing step”

• Data Integrity/Stewardship/Curation
  • “Develop a digital "chain-of-handling" to allow all relying parties to understand how, when, and by whom the results were derived. This should include code and parameters to aid in scientific reproducibility and post facto validation”
  • “…if data are not adequately curated, it is questionable whether they should have been generated in the first place”
Vision 3: Build and expand **access** to the Big Data resources and cyberinfrastructure – both domain specific and shared – that are needed for agencies to best achieve their mission goals and for the country to innovate and benefit (51%)
Education and Training

Vision 4: To educate the next generation of scientists and engineers and fulfill increasing demand for analytical talent and capacity for the broader workforce. (27%)

- Develop PhD fellowships for students who excel in either the foundations or exploitations of data science; allow longer time to joining faculty or industry to learn both technology and data in different disciplines
- Encourage public programs and certifications for roles at data-driven organizations, government competitions and prizes
- Leverage online learning
- Create hubs to enable curricula development
- Create graduate fellowships to add analytical skills
- Provide domain relevant data/software for teaching materials
- Redesign high school and collegiate science curriculum to incorporate data analytics
New Partnerships

*Vision 5: Foster the creation of new partnerships that cross sectors and domains (43%)*

- Encourage **inter-laboratory and cross-divisional collaboration**
- Fuse government and **social media data** to draw intelligence
- Work with **Nat’l Research Council** to conduct unbiased assessment on strategic scientific data investments
- Partner with **independent consulting firms** to handle data policy compliance
- Devise **industry incentives** to develop solutions
- Create **hubs** to foster public-private collaborations
- Invest in the **open source** community
- Commercial partnerships for foundational software
Data Sustainability

Vision 6: “Ensuring the long term sustainability, access, and development of high value data sets and data resources” (21%)

• Long-term investments required for all aspects of cyber-infrastructure
  • New/indefinite funding models needed
• Stability and reliability through transitions
• Integration with library and info science communities
• Formation of new cross-agency entity
• Consideration of costs of archival of all data
• Consolidate data lifecycle management across agency resources
Gateways

Vision 7: “Creating new gateways that enable the interconnection and interplay of Big Data ideas and capabilities across agency missions” (10%)

• Portals for sharing large data sets and analysis tools
• Scalable, API-driven software services
• Diversity of software packages and stacks
• Built with end-user in mind
  • Hide underlying infrastructure
  • Accessible with little/no programming experience
• Implementation in conjunction with data centers
Gaps

What are the gaps that are not addressed in the Visions and Priority Actions Document? (39%)

- Organization/Mobility
- Data Bottleneck
- Long-term sustainability
- Energy of storage, processing and transmission.
- Metadata – Argonne National Lab
- Pace of analysis tool development
- Economic value
- Privacy
- Data Migration
Game-changers

What new research, education and/or infrastructure investments do you think will be game-changing for the big data innovation ecosystem? (21%)

• Use of J/TB (Joule per Terabyte) as a metric
• Cost of data storage
• Disaster Forecasting, Health sensors/records, education tracking
• Fusion of government and social media
• Big Data Research Centers
• Incentivization
• Big Data Infrastructure (on par with supercomputing centers)
1. Data Governance is needed for a viable Big Data ecosystem

2. New definitions of risk, harm and disclosure will be needed

3. Peoples rights to negotiate the terms of their data usage
Response Coordination Team

- NSF AAAS Fellows:
  - Renata Rawlings-Goss
  - Alejandro Suarez
  - Elizabeth Burrows
  - Lida Beninson
  - Martin Wiener

- NSF Staff
  - Fen Zhao
  - Dana Hunter

- NITRD Staff
  - Wendy Wigen