The NSF Convergence Accelerator Pilot
Track A – Open Knowledge Network

Chaitan Baru, Senior Science Advisor
Convergence Accelerator Office, Office of the Director
National Science Foundation
Outline

• The NSF Convergence Accelerator Pilot program
• Convergence Accelerator Pilot Tracks
• CA Track A: Open Knowledge Network
• OKN: Projects and topics
• Next steps, Future Directions
NSF Big Ideas

RESEARCH IDEAS

- Harnessing Data for 21st Century Science and Engineering
- Work at the Human-Technology Frontier: Shaping the Future
- Windows on the Universe: The Era of Multi-messenger Astrophysics
- The Quantum Leap: Leading the Next Quantum Revolution
- Understanding the Rules of Life: Predicting Phenotype
- Navigating the Now Arctic

PROCESS IDEAS

- Mid-scale Research Infrastructure
- NSF 2050
- Growing Convergent Research at NSF
- NSF INCLUDES: Enhancing STEM through Diversity and Inclusion
The NSF Convergence Accelerator Pilot Program: Convergence Research

The **convergence paradigm** intentionally brings together intellectually-diverse researchers to develop effective ways of communicating across disciplines by adopting common frameworks and a new scientific language…

To solve the grand challenges of today that require the merging of ideas, approaches and technologies from widely diverse fields of knowledge to stimulate innovation and discovery.
**NSF Convergence Accelerator**

**WHY:** Leverage the science across all fields of NSF research to produce outcomes in an accelerated timeframe, with streamlined operations allowing for nimbleness to support the most innovative results

**WHAT:** A new organizational structure to *accelerate* the transition of convergence research into practice, in areas of national importance

### Characteristics

- Use-inspired research
- Testbeds, tools, living labs...
- Larger, national scale
- Requires partnerships with industry
- Clear goals, milestones, directed deliverables

### Management

- Time-limited “tracks”
- Teams and Cohorts
- Cooperation and Competition
- More directed management
- Mission-driven evaluation
Convergence Accelerator Pilot

Tracks

- **Harnessing the Data Revolution and Future of Work at the Human-Technology Frontier**, selected as the track themes for the Pilot
- **Open Knowledge Network and AI and Future Work**, selected as the track topics

Choosing track topics

- Use-inspired research; require convergence, i.e. strong multidisciplinarity; national priority areas; doable within the NSF CA format (6 months planning; 2 years of implementation)

Research Concept Outlines (RCOs)

- Community is invited to submit 2-page RCOs about their track ideas.
- RCOs are reviewed, and teams are invited to submit Phase I planning proposals based on the RCOs.
Tracks for 2019 CA Pilot

**Track A1**
**Goal:** Enhancing scientific data discovery and use
**Track:** Open Knowledge Networks
**Big Idea:** Harnessing the Data Revolution

**Track B1**
**Goal:** Connecting, retraining and reskilling for jobs using AI
**Track:** AI & Future Jobs
**Big Idea:** Future of Work at the Human Technology Frontier

**Track B2**
**Goal:** Building STEM talent in a changing workplace
**Track:** National Talent Ecosystem
**Big Idea:** Future of Work at the Human Technology Frontier

**Vertical:** Challenges specific to different topical domains such as geosciences, education, smart health, finance, and manufacturing.

**Horizontal:** Challenges that apply to all domains, such as developing the underlying representation of facts or developing secured access capabilities.

Harnessing the Data Revolution

Future of Work at the Human-Technology Frontier
CA Track A: Open Knowledge Network

• Responds to the need identified in the NSF Harnessing the Data Revolution Big Idea for:
  • An advanced science data infrastructure that is interoperable and has an open architecture, making it easier to access and link heterogeneous data products
  • An open semantic information infrastructure to discover new knowledge from multiple disparate knowledge sources
  • A nonproprietary shared knowledge infrastructure, with a particular focus on publicly available data, e.g., U.S. Government, scientific data, and other similar public datasets.
• Benefits multiple applications domains, including science and engineering research
The Open Knowledge Network Idea

• Jan 2015: Big Data Strategic Initiative workshop, Georgetown Univ.
  • Prof. Andrew Moore describes how industry is using knowledge networks as key enabling technologies, and the need for “reference entities”.

• June 2016: Meeting at White House OSTP
  • Small group with govt, industry, academia

• Feb 2017: Meeting in San Jose, hosted by AWS
  • ~55 attendees; ~40 companies
  • Lots of interest in the idea

• Oct 2017: NITRD OKN Workshop
  • Enabling the Community to Build the Network, Oct 4-5, 2017, National Library of Medicine

OKN: Motivation

- **Natural interfaces** to large knowledge structures have the potential to impact science, education and business to an extent comparable to the WWW.
- The first wave has appeared in consumer services, e.g., Siri, Cortana, Alexa.
  - But limited in their scope to specific business areas, and proprietary, not open
- An open effort could help expand to 1000’s of new topic areas, and many more useful classes of questions—even with current technologies.
  - A “Siri for science”
- Requires convergence among technology areas and domain sciences


- NSF has supported significant past research in
  - Creation of knowledge bases—representation, performance, creation of ontologies, knowledge extraction, knowledge aggregation, reasoning ...
OKN Phase 1 Projects

“Vertical” Projects
- A7160 Precision Medicine
- A7043 Design & Manufacturing
- A7099 Urban Flooding
- A7152 Space Sciences
- A7017 Molecular Data
- A7095 Census
- A7123 Court Records
- A7115 Civil Infrastructure
- A7137 Energy Systems
- A7134 Intelligent Textbooks
- A7033 Public Policy Data
- A7123 Court Records
- A7095 Census
- A7153 Finance
- A684 Mobility
- A6950 Ocean Resources

“Horizontal” Projects
- A6940 Knowledge Application Development Environment
- A7165 Internet Structure & Security
- A6731 Web Data Extraction/Integration
- A6677 Spatial Data Models/Methods
- A7908 Spatial Decision Support
- A7143 Information Credibility
- A7136 Federated Search

Projects should
- Seek “track integration”;
- Collaborate with industry;
- Strive to collaborate/link with other relevant efforts in the community
OKN: Some common themes across projects

• Integrating heterogeneous types of data
  • E.g., distinct types of biomedical data: biomedical facts; bioinformatics catalogs of data; “raw” data from measurements;
  • E.g., internet configuration and traffic data at various levels in the network

• Accommodating dynamic information
  • Newly acquired data and information are related in a network fashion to extant knowledge
  • Need the capability to add new vertices and edges as new knowledge is added and schemas change
OKN: Common themes...

• Supporting access by and contributions to the KG by heterogeneous communities of users
  • Due to the multidisciplinary nature of the problems being addressed, experts as well as non-experts from each related domain must be able to contribute and share data and knowledge, and access the network to obtain answers to questions beyond their own area of expertise
• Incorporating new information into the knowledge graphs using ML + crowdsourcing approaches
  • E.g., Intelligent TextBooks powered by knowledge graphs. Scaling out to many domains--capturing relationships among concepts in a domain using experts(crowdsourcing) as well as ML techniques
• Several Track B projects also planning to develop and refine ontologies related to job skills
Next Steps

• CA Phase 1: Planning (Sept 2019 – May 2020)
  • Monthly meetings with required participation in *Convergence Accelerator curriculum*
    • Design thinking and user-centered design
    • Team Science
    • Domain-specific talks and interactions with potential collaborators
  • Work with team coaches
  • Team refinement
  • User interviews: Must conduct at least 12 user interviews
CA Phase I...

• Monthly meetings with the full cohort (43 teams x 3 per team)
  • September 2019:  Webinar
  • October 2019:  Kickoff in Washington DC. Interaction with government agencies.
  • November 2019:  Webinar
  • December 2019:  Face-to-face in San Francisco. Meetings at IDEO (design thinking). Interactions with industry.
  • January 2020:  Webinar
  • February 2020:  Face-to-face in San Francisco. Interact with foundations, VCs, industry.

• Projects are also expected to organize community workshops to obtain input
CA Phase II: Implementation

• Implementation Phase: 1+1 years (June 2020 – June 2022)
• April 2020: Submit Phase II proposal
• May 2020: Make a ~8-10 minutes “pitch” to a CA review panel and other potential co-funders including Foundations, VCs, other agencies, other NSF programs, other investors...
• Projects will likely be evaluated on the submitted proposal; their pitch; and “expressions of interest” from potential co-funders
• Total ~5 awards per Track of up to $5M each (up to $3M in Year 1)...or more, based on co-funding
  • Cooperative agreements
  • Year 2 funding based on Year 1 performance
  • Need to demonstrate progress on quarterly milestones
Next Steps / Future Directions

• OKN Projects are encouraged to link with other related efforts in the community, including industry
  • Includes NSF AI Institutes; NIH, DOE, NIST, etc efforts in AI
• New CA tracks under consideration for 2021, including any “gap filling” that may be needed for Tracks A and B
  • Workshops have been conducted in: Quantum computing; Carbon Sequestration; AI and Disasters; Data Privacy and Data Governance; Innovative use of data science tools with scientific data, for research and education, ...
• Sustainability of efforts
• Strategic directions for the CA program
CA Timeline – 2019 Pilot and the Future

2019 Pilot Cohort (Tracks A and B)

- Accelerator
  - DCL issued
  - RCOs

Phase 1: Team formation, res. plan development

Phase 1: Project Proposals Start

Phase 1: Innovation Curriculum

Pitch Competition

Phase 2: Creating Deliverables

Phase 2: Projects Start

Year 2: Decision

Deliverables

2020 Cohort: New tracks (C and D)

- RFI: 75 responses submitted
  - 2020 Topic Workshops

2020 Solicitation

2020 RCOs

Phase 1: Project Proposals Start

Innovation Curriculum

Pitch Competition

Phase 2: Projects Start

Year 2: Decision

RCOs

Ontology Summit, Feb 5, 2020
Accelerator “Tracks B1 and B2”: FUTURE OF WORK AT THE HUMAN-TECHNOLOGY FRONTIER

• **AI and Future Jobs.** The AI and Future of Jobs track will support the development of mechanisms that connect workers with jobs of the future, reflecting the need for re-skilling and lifelong learning, such as predictive artificial intelligence tools, economic and labor market analyses of needed skills for future workplaces, and educational technologies needed for adult learning. Ensuring fair and ethical treatment of workers will be a key principle for this effort. Projects may be focused on particular industries or regions, specific populations such as veterans, or particular workplace types such as small businesses, manufacturing, or K-12 schools.

• **National Talent Ecosystem.** Innovative approaches for employers to support workers seeking the skills required for 21st century work related to data science, predictive analytics, AI/machine learning, and other technologies of the future. Successful projects will prototype innovative approaches, such as learning environments, simulations and tools for analysis or assessment, and vehicles for recruitment and engagement, with the potential for wider implementation by industry, educational institutions, and other stakeholders engaging in the co-creation of a national talent ecosystem.
Track B1/B2 Projects

B7026 – Machine learning-based national labor market information tools
B6970 – AI+AR platform for autism spectrum disorder workers
B6857 – AI-based job matching – veterans, disabled workers
B7068 – Documents competencies at the national level
B7118 – Connects data exchanges at state level

**Worker-Work Matching**

**Existing Qualifications:**
- Education
- Skills
- Certificates

**Prospective Employee**

**Prospective Employer**

**Eduction/Training**

**Existing needs:**
- Positions
- Skill requirements
- Locations

**Future needs:**
- Emerging jobs
- Growth projections
- Market demands

**Workforce Training and Education Recommendations**

- B6947 – National microcredential system
- B7063 – Microcredential system for industrial robotics technicians
- B6992 – AI-enabled assessment + training plan for displaced miners
- B7037 – AI-driven skill gap diagnostics + recommendation engine for manufacturing
- B7010 – Assessment/Prediction/Learning – smart sensing/mixed reality
- B6968 – Machine learning based tools for gig economy workers
- B6956 – AI-driven tool for career management in STEM fields
- B7888 – Fostering a diverse AI workforce

**Curricula and Skills Training Development**

- B6894 – Upskilling/reskilling for digital technologies
- B6656 – Design based research + analytics identifies skill gaps and designs training
- B7833 – Deep learning predicts future jobs + training for hospitality industry
- B6997 – Training platform for autonomous systems
- B7063 – Advanced robotics for training next gen emergency responders
- B7019 – Cloud-based platform trains for future jobs in architecture, construction
- B7061 – Develops ROI measurement for training programs for policymakers
- B7130 – Low cost AR training content development platform for SMEs

**Existing needs:**

- Positions
- Skill requirements
- Locations

- Emerging jobs
- Growth projections
- Market demands

**Education/Training**

Ontology Summit, Feb 5, 2020

- Up to $1M for ~9 months: Planning, team formation, required participation in Convergence Accelerator (CA) curriculum
  - Design thinking and user-centered design. Provided by IDEO.
  - Team Science
  - Domain-specific talks and interactions with potential collaborators
- Each team assigned a coach from a team of coaches
  - Must conduct at least 12 user interviews
  - May consult with any of the coaches, if they wish.