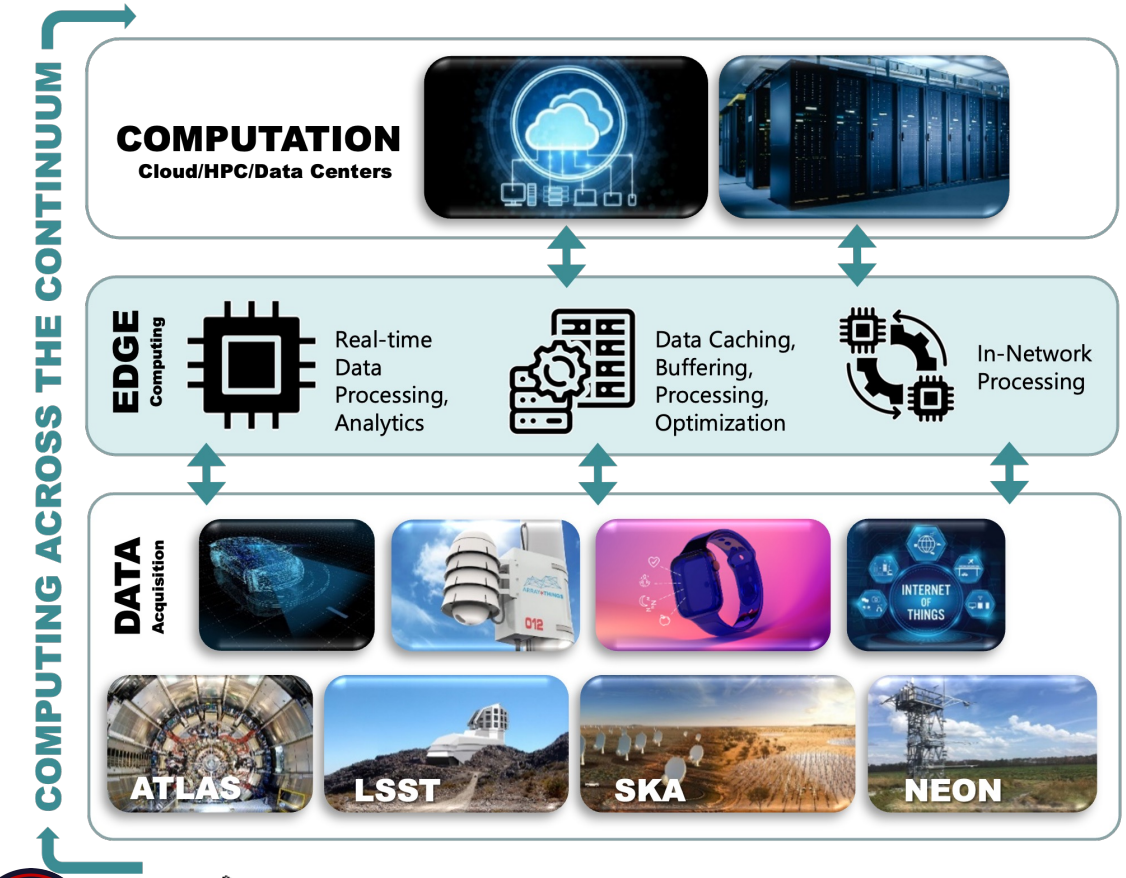
Facilitating trade-off management on the Edge-Cloud Computing Continuum for Urgent Science

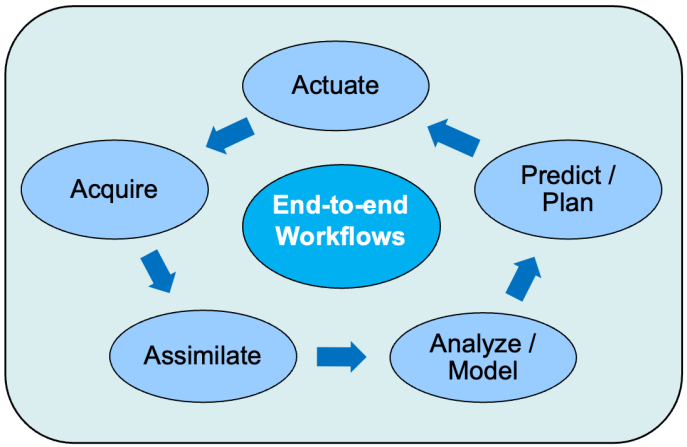
Manish Parashar (Utah) & Daniel Balouek (Inria)

2025 InPEx Workshop
Kanagawa, Japan, April 15-17, 2025

Scientific Computing and Imaging (SCI) Institute
One-U Responsible AI Initiative

A Digital Continuum for Science





Manish Parashar. 2024. Everywhere & Nowhere: Envisioning a Computing Continuum for Science. https://doi.org/10.48550/arXiv.2406.04480.

Daniel Balouek-Thomert, Ivan Rodero, and Manish Parashar. 2020. Harnessing the Computing Continuum for Urgent Science. SIGMETRICS Perform. Eval. Rev. 48, 2 (September 2020), 41–46. https://doi.org/10.1145/3439602.3439618

Global Challenges and Urgent Science

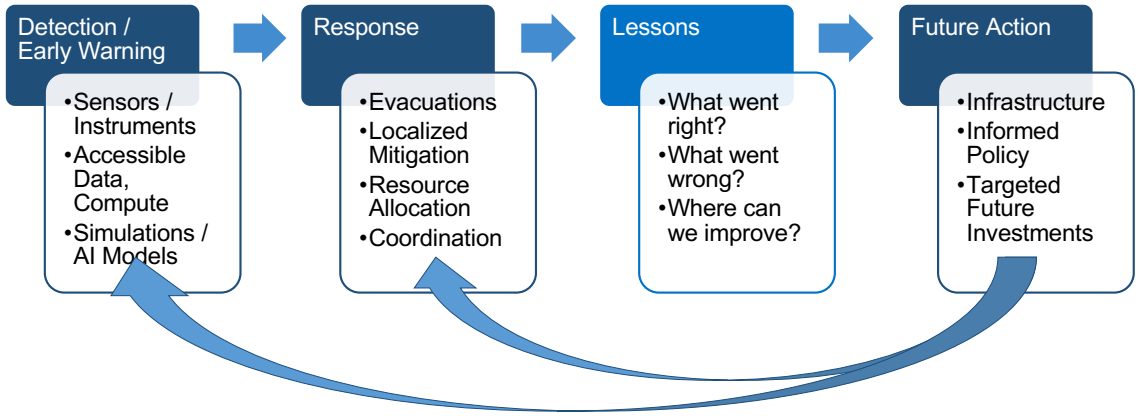
Earthquakes & Tsunamis

Extreme Weather

Cyber-Attacks

Pandemics

Industrial Disasters



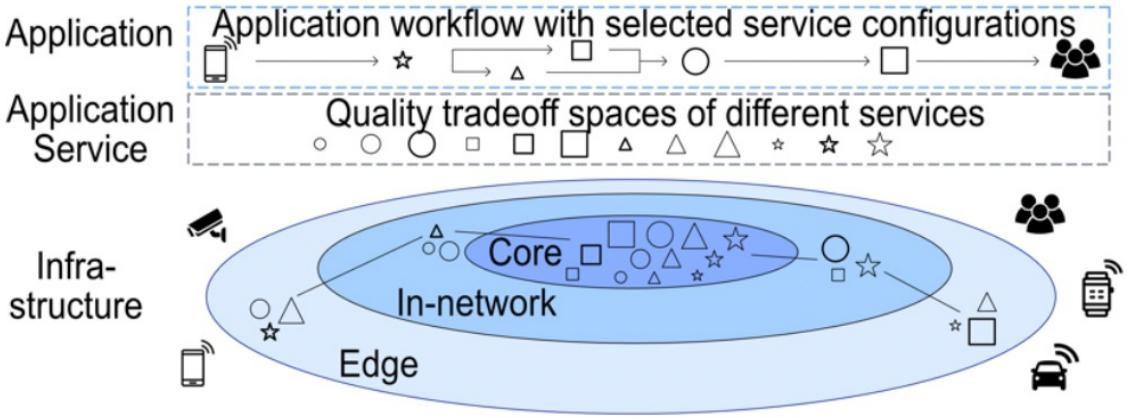
Urgent Computing: Managing QoS under constraints, uncertainty

DEFINITION: Computing under **strict time** and **quality constraints** to support decision making with the desired confidence within a defined time interval

Urgent applications need to react to unforeseen events and to manage complex cost/benefit tradeoffs to meet constraints

- Steer online/on-demand computations based on data/content
- Support urgency constraints for timely decision-making
- Balance “costs” of computations with value added, availability of resources, etc.

Data-driven Urgent Application Workflows



- End-to-end application workflows implemented as compositions of services spanning data producers located at the edges, in-network resources, and data consumers at the core.
- New data trends trigger reconfigurations of the application workflow according to available resources.

- Possible information flows:
  - Pull: data generated in an energy-constrained environment and consumed in an energy-rich environment
  - Push: data generated in an energy-rich environment and consumed in an energy-constrained environment
  - Push-Pull: data is produced and consumed at both ends

Urgent Computing: Many Research Challenges

- H
  a
- H
- H
  e
- H
- H
  o
- H
  o
- H
- H
- V
  e
- **…**

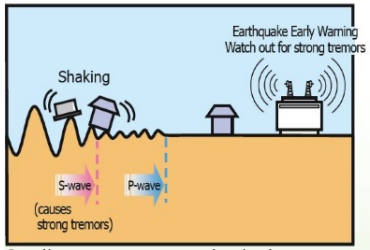
Managing tradeoffs: How to manage performance/cost tradeoffs across the computing continuum to (1) understand the variability of the resources composing the continuum for service placement and orchestration and (2) enable control of network components through collaborative learning.

- Where and when to run training processes considering computation, communication capacity, and cost?
- How to detect anomalies and provision resources?

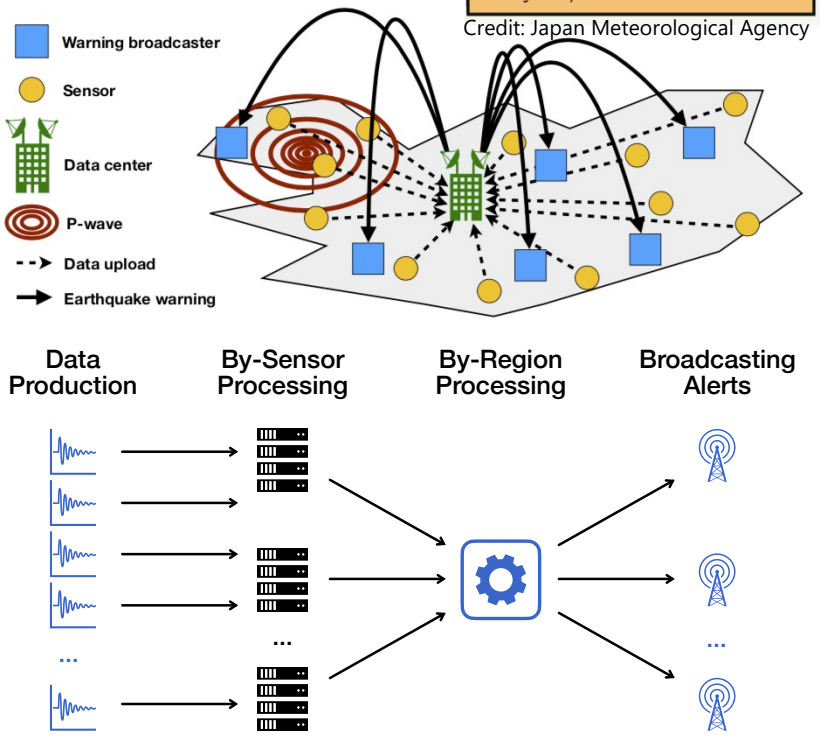
Earthquake Early Warning

- **Earthquake Early Warning (EEW)** requires earthquakes to first be characterized (magnitude, location, speed of displacement, etc.)
  - A ***single data source*** doesn't able to cover a whole spectrum of events:
    - Seismometers are good for the smaller earthquakes (< 6.5); High-precision GPS are good for larger earthquakes.
  - ***Centralized data processing*** cannot support real-time, high volume data processing

- **Goal:** Combine ***multiple data sources*** to cover the whole spectrum of events; leverage the ***CI continuum*** for preserving latency and resiliency



Credit: Japan Meteorological Agency



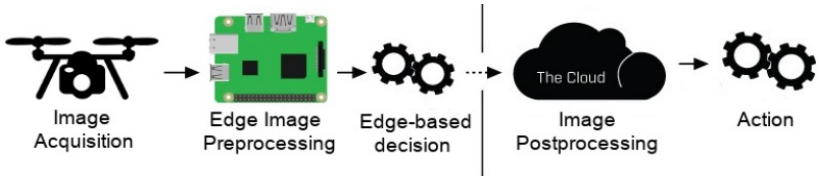
Fauvel, K, et al. "A distributed multi-sensor machine learning approach to earthquake early warning," Proceedings of the AAAI 2020, **Outstanding Paper Award in Artificial Intelligence for Social Impact**.

Hurricane Sandy -- Disaster Response

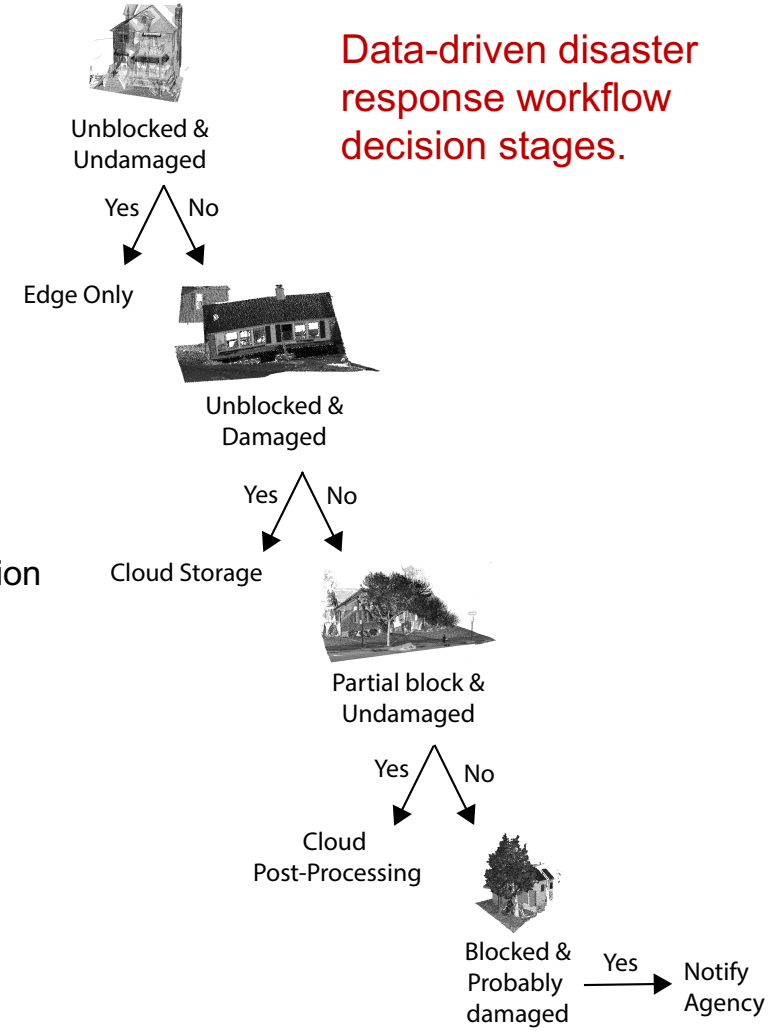
Goal: Quickly and efficiently determine whether the civil infrastructure (buildings, bridges) conditions are safe or not for evacuees to return

Edge Approximation: Canny-edge detection algorithm

- Multi-stage algorithm to detect a wide range of edges in images
- Ease to apply approximation techniques (substitution, discarding) at function and input parameters level



E. Renart, D .Balouek-Thomert, X.Hu, J. Gong and M. Parashar. Online Decision-Making Using Edge Resources for Content-Driven Stream Processing, eScience'17, 2017



Data-driven disaster response workflow decision stages.

A Motivating Usecase: Wildfire in California leads to Air Pollution in Utah

California's fast-moving Oak Fire burns 14,000 acres and forces thousands to evacuate outside Yosemite National Park

By Jason Hanna, Rebekah Riess, Sara Smart and Andy Rose, CNN

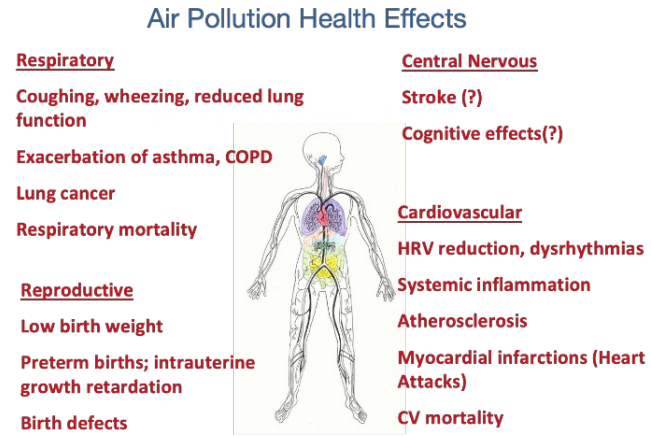
Updated 0645 GMT (1445 HKT) July 25, 2022

Could the exception become the rule? 'Uncontrollable' air pollution events in the US due to wildland fires

Liji M David[1,2], A R Ravishankara[1,2], Steven J Brey[2], Emily V Fischer[2], John Volckens[3] and Sonia Kreidenweis[2]

Published 22 February 2021 • © 2021 The Author(s). Published by IOP Publishing Ltd

Detection ➡ Emission ➡ Transport

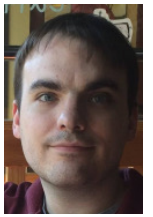


Mitigating the negative impacts of wildfires on air quality requires combining knowledge from multiple data sources and integrating it on-demand with distributed computational models.

Balouek-Thomert, D., Caron, E., Lefèvre, L. & Parashar, M., Towards a methodology for building dynamic urgent applications on continuum computing platforms, Combined International Workshop on Interactive Urgent Supercomputing (CIW-IUS), 11/2022.

Implementing such urgent workflows can be hard: Technical challenges include

- How to **drive** computation through data?
  - R-Pulsar programming system of data-driven workflows (doi: 10.1109/UrgentHPC54802.2021.00007)
- How to **discover** and aggregate resources based on current needs?
  - Software defined system driven by constraints (doi:10.1177/1094342017710706)
  - Science recommender systems for intelligent data discovery (doi.org/10.1109/MCSE.2022.3179408)
- How do you **manage** execution (QoS) in a dynamic environment?
  - Autonomic runtime management (doi.org/10.1109/MIC.2020.3039551)
  - Virtual data collaboratory/National Data Platform (doi: 10.1109/MCSE.2019.2908850; 10.1109/e-Science58273.2023.10254930)


P. Davis

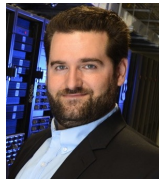

D. Balouek-Thomert


C. Sicari


M. AbdelBaky


Y. Qin


S. Alhariri


I. Rodero


M. Adair

National Data Platform: Customized Data Experiences through Near Data Services

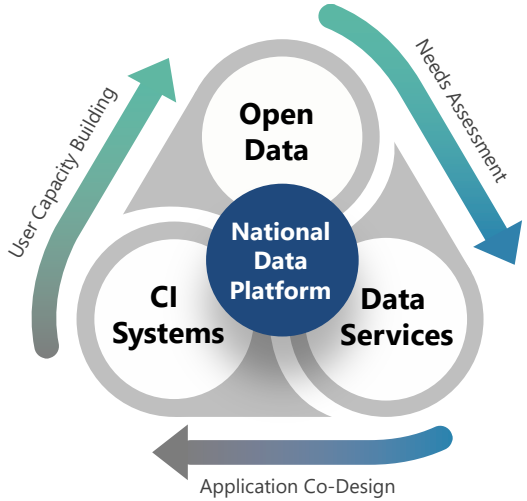
#2333609

A **federated** and **extensible** data ecosystem to promote innovation and collaboration through the equitable use of data leveraging existing and future national cyberinfrastructure capabilities.

https://www.nationaldataplatform.org/

FOCUS AREAS:

- **Platform** for data-enabled and AI-integrated workflows
  - Facilitates data registration and discovery via a **centralized hub**
  - Democratizes data access and use via **distributed points of presence**
  - Cultivates resources for **classroom education** and **data challenges**
  - Assists research and learning through **personalized workspaces**



- **Applications** in climate and AI with data diverse scientific data repositories including NSF facilities, NAIRR, NASA, USGS, NOAA and USDA

- **Partnerships** to foster scientific discovery, decision-making, policy formation and societal impact

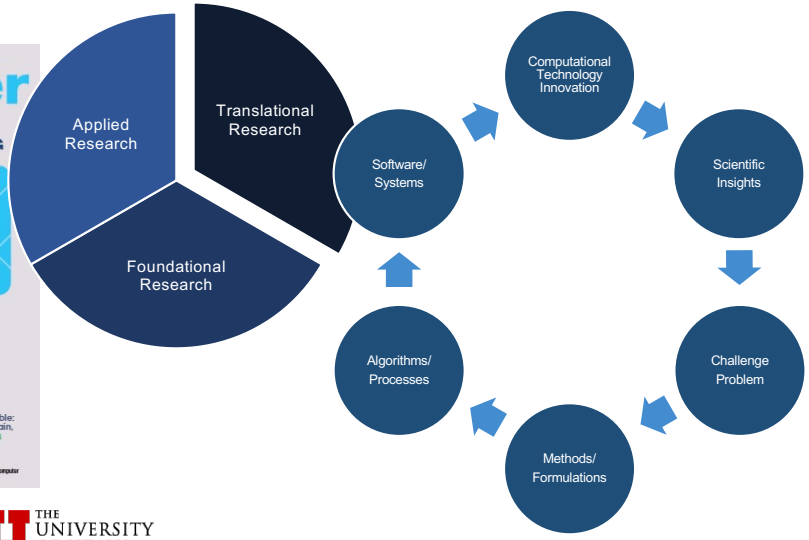
UC San Diego · The University of Utah · University of Colorado Boulder · San Diego Supercomputer Center · SCI www.sci.utah.edu · EarthScope Consortium

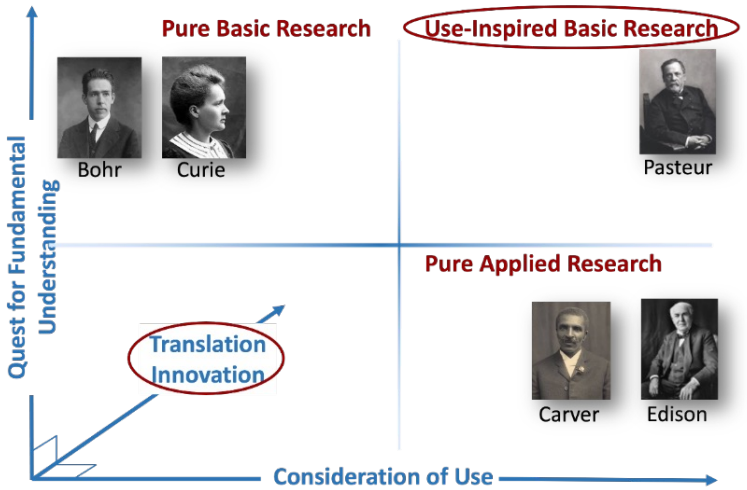
Translational Research and Innovation

translational-cs.org

Translational Research refers to research that bridges foundational, use-inspired, and applied research with the *delivery and deployment of its outcomes* to the target community and supports essential *bi-direction interplays* where delivery and deployment process informs and advances research

- Taking research from the *Laboratory* to the *Locale* to the *Community*





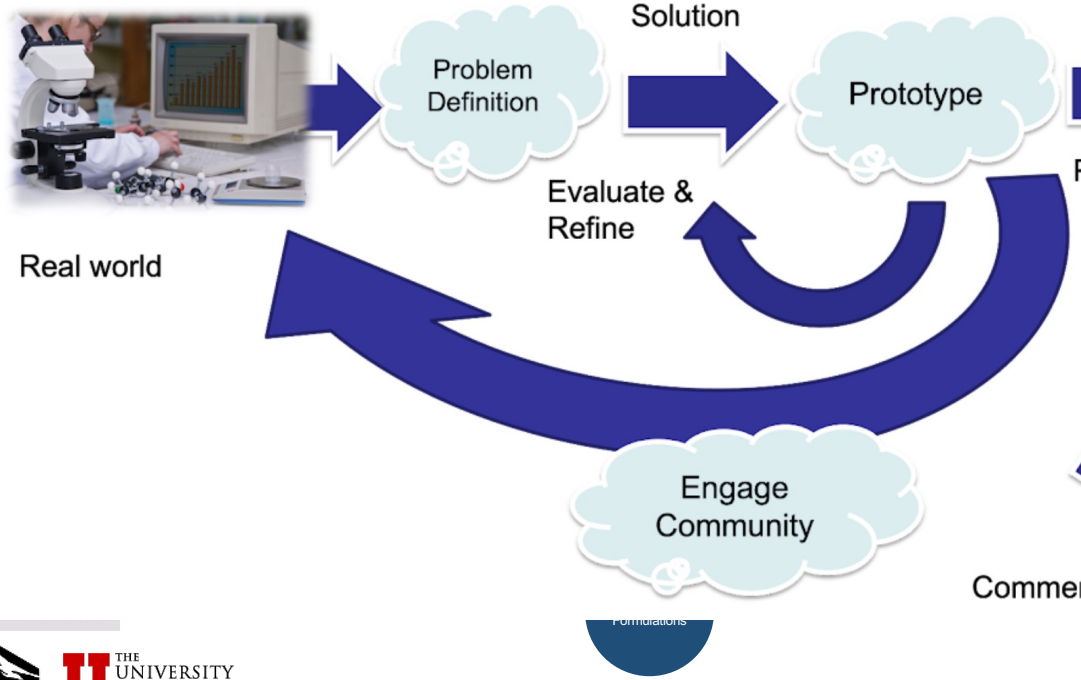
Stokes, Donald E. (1997). Pasteur's Quadrant – Basic Science and Technological Innovation. Brookings Institution Press. p. 196. ISBN 9780815781776. https://archive.org/details/pasteursquadrant00stok/page/n9/mode/2up

Translational Research and Innovation

translational-cs.org

Translatic
foundation
and deplo
supports e
deploymer

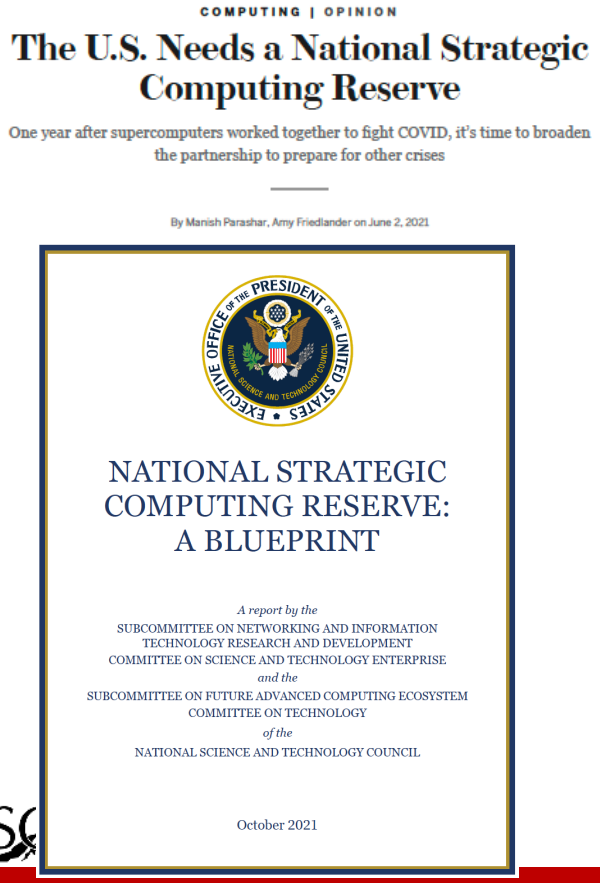
- Taking re



c Science
n Press.

:/page/n9/

National Strategic Computing Reserve (NSCR)



COMPUTING | OPINION

The U.S. Needs a National Strategic Computing Reserve

One year after supercomputers worked together to fight COVID, it's time to broaden the partnership to prepare for other crises

By Manish Parashar, Amy Friedlander on June 2, 2021

EXECUTIVE OFFICE OF THE PRESIDENT OF THE UNITED STATES
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

NATIONAL STRATEGIC COMPUTING RESERVE: A BLUEPRINT

A report by the
SUBCOMMITTEE ON NETWORKING AND INFORMATION TECHNOLOGY RESEARCH AND DEVELOPMENT
COMMITTEE ON SCIENCE AND TECHNOLOGY ENTERPRISE
and the
SUBCOMMITTEE ON FUTURE ADVANCED COMPUTING ECOSYSTEM
COMMITTEE ON TECHNOLOGY
of the
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

October 2021

- Advanced computing cyberinfrastructure can be a strategic National asset in emergency response, if mobilized quickly.

  Goals for an NSCR:

  - Ensure availability of a ready "reserve" of resources (computing, data, software, services) and expertise that can be leveraged nimbly in times of urgent need.
  - Establish policies, processes, and agreements to enable agile, effective, and impactful resource mobilization.
  - Build on continued longer-term strategic investments in resources (computing, data, software/services), expertise.
  - Coordinate across agencies, stakeholder communities, and other national reserves.

Full report available at:
https://www.whitehouse.gov/wp-content/uploads/2021/10/National-Strategic-Computing-Reserve-Blueprint-Oct2021.pdf

Thank you!

One-U RAI Opportunities

- **Distinguished Visitors Program:** Supports visits from a few days to a full year for faculty, up to two years for postdoctoral fellows.
- **Postdoctoral Fellows Program:** Supports postdocs in areas related to responsible AI for up to 2 years.

Manish Parashar

Email: manish.parashar@utah.edu

WWW: manishparashar.org / sci.utah.edu / rai.utah.edu

one-u-responsible-ai-initiative

@OneU_RAI

SCI Institute

Transformation of science and society through ***translational research and innovation***

- Inter/transdisciplinary, collaborative, convergent
- Core strengths in: Visualization & imaging; Scalable analytics; Advanced computing & data
- Software/system development and distribution integral to our research processes