



**NORTHEAST
BIG DATA**
INNOVATION HUB

**BUILDING MULTI-SECTOR PARTNERSHIPS
TO ADDRESS SOCIETAL CHALLENGES WITH
DATA-DRIVEN SOLUTIONS.**

INTRODUCTION TO THE BIG DATA HUBS

2012

Launch of the National Big Data Research and Development Initiative.

2013

Mission: Solving pressing R&D challenges related to extracting knowledge and insights from large and complex collections of digital data.

2014

Strength in numbers: This initiative encouraged multiple stakeholders to develop and participate in Big Data research and innovation projects across the country.

2015

Launch of National Network of Big Data Regional Innovation Hubs (Big Data Hubs).

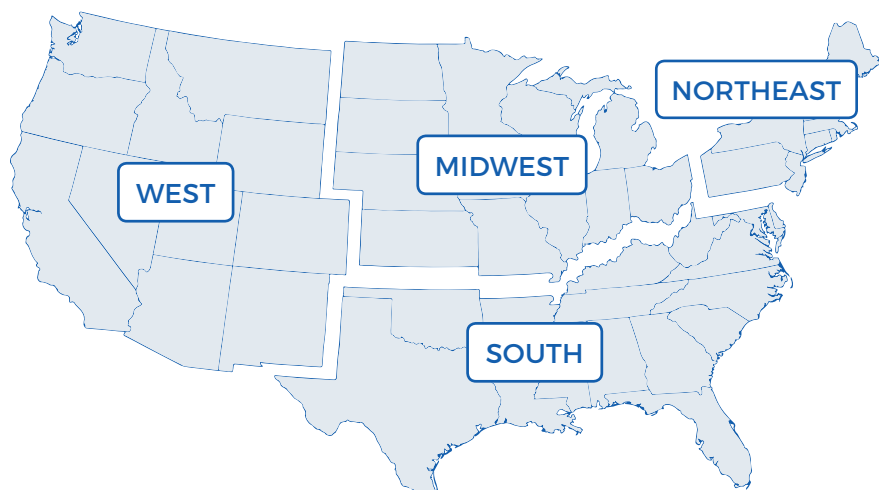
Established by the National Science Foundation's Directorate for Computer and Information Science and Engineering (CISE).

What is a “Big Data Hub”?

Each Big Data Hub is a uniquely neutral platform that supports its constituent members—drawn from academia, industry, non-profit organizations, and/or government—to work in concert to achieve common Big Data goals that would not be possible for the independent members to achieve alone. Priority areas and projects are regionally defined, addressing Big Data challenges and opportunities.

The Big Data Hubs represent the four Census Regions of the United States.

Alaska and Hawaii are part of the West Region. US Territories can participate in any region.



INTRODUCTION TO THE NORTHEAST HUB

The Northeast Big Data Innovation Hub builds public-private, multi-sector partnerships, and consortia to address high-priority challenges with data-driven solutions.

Our founding Priority Areas.

Cities & Regions

Using big-data analytics to generate operational efficiencies and improve quality of life in the highly-urban Northeast's cities and surrounding regions.

Data Literacy

Identifying knowledge and resource gaps, with a goal to help lifelong learners of all ages become data literate, throughout the Northeast and beyond.

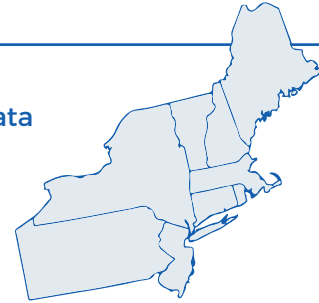
Data Sharing

Addressing key data sharing challenges relating to policy, privacy, platforms and formats, software, costs, and promoting data sharing benefits.

Discovery Science

Uncovering new knowledge through data driven modeling and discovery; ensuring repeatability, curation and provenance of massive datasets and computational tools and practices.

We represent data science across sectors in the Northeast.



Education

Employing data-driven approaches to education; pairing deep knowledge of learning and cognition with adaptive curricula; and enabling innovative instructional resources.

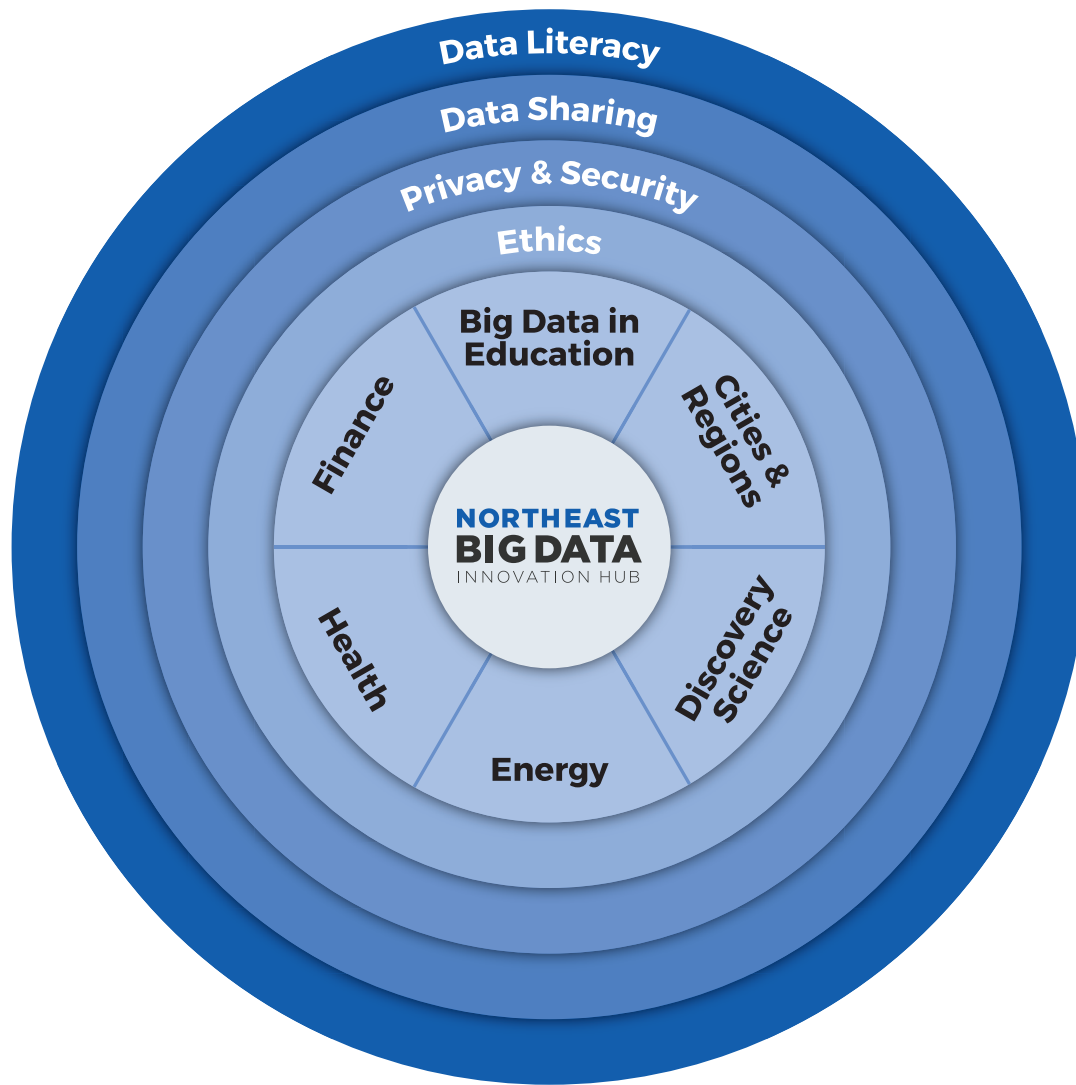
Energy

Analyzing and effectively using data from all phases of the energy cycle to more effectively model the energy ecosystem and drive innovation.

Ethics

Anticipating issues and unforeseen consequences of Big Data technologies; offering practitioners and regulators insight and direction to address them; making the most out of Big Data's potential to benefit society and minimizing its potential harm.

Continued »



Finance

Leveraging regional strengths in finance and economics; employing data-driven approaches to improve our understanding, modeling, and management of modern markets.

Health

Acquiring and integrating data and knowledge; advancing analytics, including causal discovery and reasoning, AI, and machine learning; and enhancing precision medicine, including advanced decision support.

Privacy & Security

Advancing the Big Data agenda on privacy and security; building a community of interested researchers and practitioners; and catalyzing regional activities.

The following pages will focus on the three projects and four planning projects which received NSF funding in the initial call for “Big Data Spokes” (priority areas). More information on all of our priority areas is available on our website.

DATA SHARING

Facilitating the exchange of solutions to advance data sharing among a range of stakeholders.

As a community, we seek to address key data sharing challenges relating to policy and privacy, platforms and formats, software and costs, and ethics and education about data sharing benefits.

The NSF Big Data Spoke project, “A Licensing Model and Ecosystem for Data Sharing,” is addressing some of these challenges. Our team is developing a safe and secure data sharing platform that facilitates sharing data that may or may not be open or free between different organizations (industry, academia, government).

nebigdatahub.org/data-sharing



PROJECTS

BIG DATA IN EDUCATION

Creating a longitudinal educational grand challenge that uses existing data to discover factors that make a difference in students' long-term educational success.

The Northeast is a center of gravity for innovations in education, anchored by universities and publishers who drive K-12 education in America. This project will improve capacity in data-driven education by sharing educational databases, managing yearly data competitions, and conducting educational data science workshops and hackathons. The team intends to improve classroom learning and leverage the unique types of data available from digital education to better understand students, groups and the settings in which they learn.

nebigdatahub.org/education

A close-up photograph of a person wearing a white lab coat and blue nitrile gloves. They are holding a silver pen and pointing at a laptop screen. The screen displays a DNA sequencing chromatogram with multiple lanes of colored peaks (green, red, blue, black) and corresponding nucleotide sequences (A, T, C, G) below them. The background is a blurred laboratory setting with various glassware and equipment.

HEALTH

Assembling a first-ever data warehouse to house numerous health/clinical, environmental, behavioral, and economic data streams.

By breaking current data silos and bringing together multiple large environmental and clinical data streams, this project will enhance health research, allowing causal discovery between these data sources. The ultimate goal of the project is to facilitate community-led and collaborative causal discovery through dissemination of integrated and open big data and analytics tools.

nebigdatahub.org/health

PLANNING PROJECTS



PRIVACY & SECURITY

Advancing the Big Data research agenda on privacy and security.

No matter the domain area, issues fundamental to big data—individual privacy, protection of data, data authenticity, and rights such as free speech—must be addressed.

Two planning projects in Privacy & Security:

“Planning for Privacy and Security in Big Data.” Bringing together stakeholder communities to:

- › Understand how privacy currently limits data sharing.
- › Develop standards and best practices to enable new information flows.
- › Highlight privacy and security issues associated with our priority areas.

“Cross-Organization Big Data Cyber Attack Awareness.”

- › Focus on protected sharing of cybersecurity data for countering attacks on digital infrastructure.
- › Developing a platform to enhance collaborative cyber security operations through cross-organization sharing of relevant cybersecurity data.

nebigdatahub.org/privacy-security



DATA LITERACY

Charting a path forward that bridges big data practice with big data learning, education and career readiness.

Data-driven science and technology pervade 21st century life and work. Yet there has been little effort spent on developing resources or frameworks to support the kind of literacies necessary for a diverse populace to understand, engage with, and leverage big data for use by themselves and their communities. To address these challenges, we are actively identifying knowledge and resource gaps, with a goal to help lifelong learners of all ages become data literate, throughout the Northeast and beyond.

nebigdatahub.org/data-literacy



ENERGY

Inspiring new projects driven by the core needs of the energy sector.

The Northeast's large industrial hubs—Pittsburgh, Buffalo, Syracuse, and others—are redeveloping and reinventing themselves, with data analytics playing a major role. The initial planning project explores how to use a brownfield redevelopment and associated energy infrastructure reinvention in Buffalo, NY as a case study to frame the energy sector's big data innovation needs.

nebigdatahub.org/energy

FULFILLING OUR MISSION

Our mission can only be realized by providing frameworks for the cross-fertilization of ideas, best practices, resources, and talent across domain areas and sectors. Beyond projects specific to our priority areas, the Northeast Big Data Innovation Hub has embarked on an ongoing dialogue with our stakeholders to understand their challenges and explore how we can leverage our networks to create programs and partnerships that help our communities thrive. These include the development of interactive events; education programs – including executive training and internships; working groups; and multi-sector projects. Some approaches we've taken:



Big Data Road Trip: Over ten cities visited in an ongoing effort to:

- › Learn more about current research/initiatives in our priority areas;
- › Discover how we can best serve our stakeholders; and
- › Learn about resources useful to our community.

Please contact us if you would like to organize a visit: contact@nebigdatahub.org



Multi-sector Consortia: The Northeast Hub has identified a number of opportunities for multi-sector consortia to substantially address challenges that no single organization or sector might solve, alone. We are evaluating whether and how we can best assist in coordinating them as a uniquely neutral organization.



Big Data Map & Assets Platform (BDMAP): We are exploring the development of a dynamic, community-driven, big data resource map which will identify, make discoverable, and visualize resources across our region and beyond.



Executive Training: We have launched an early-stage collaboration to develop an executive training pilot program to help leaders and decision-makers within companies understand how to more effectively use data science tools and teams, and how to build strategic/mission-driven data science capacity.



Innovator Internships: We matched talented data science graduate students with small- to medium-sized organizations working in our priority areas to help drive innovation where its impact would be most felt.

Connect with us at contact@nebigdatahub.org to discuss collaboration opportunities or invite the Northeast Hub to present at your events.



Here's how you can get involved:

 nebigdatahub.org

 contact@nebigdatahub.org

 [@NEBigDataHub](https://twitter.com/NEBigDataHub) | [#BDHubs](https://twitter.com/BDHubs) | [#NEBigData](https://twitter.com/NEBigData)

 [Northeast Big Data Innovation Hub](https://www.linkedin.com/company/northeast-big-data-innovation-hub)

 tinyurl.com/NEBDHubList

Northeast Big Data Innovation Hub

Room 535B Mudd Building

Mail Code 4720

500 West 120th Street

New York, NY 10027

