HARNESSING THE DATA REVOLUTION FOR 21ST-CENTURY SCIENCE AND ENGINEERING (HDR)

HDR Funding
(Dollars in Millions)

	FY 2018 Actual	FY 2019 (TBD)	FY 2020 Request
Stewardship Activities (CISE)	**-**	**-**	**$30.00**
Foundational Activities	**$169.62**	**-**	**$117.24**
BIO	3.20	-	7.80
CISE	95.63	-	55.25
EHR	4.30	-	2.50
ENG	21.58	-	12.60
GEO	2.90	-	3.00
MPS	34.62	-	20.00
SBE	7.39	-	6.09
IA	-	-	10.00
Total	**$169.62**	**-**	**$147.24**

Overview

NSF's HDR Big Idea is a national-scale activity to enable new modes of data-driven discovery in science and engineering. Access to the next level of discovery relies on translating complex data from observations, experiments, and simulations into knowledge. To help close the loop from data generation to analysis, and on to simulation and finally discovery, the HDR Big Idea will support fundamental research in data science and engineering; development of a cohesive, federated approach to the research data infrastructure needed to power this revolution; and development of a 21st-century data-capable workforce. Individuals and communities will also benefit from data-rich capabilities, infrastructure, and services.

Goals

The HDR vision is realized through an interrelated set of goals:

1. The foundations of data science: Develop the theoretical foundations of data science and its applications through integrated research and training activities;
2. Algorithms and systems for data science: Support the development and use of novel algorithms and systems to support data science as well as data-driven science and engineering;
3. Data-intensive science and engineering: Stimulate advances in multiple areas of science and engineering through data-intensive research that harnesses diverse data sources and applies new methodologies, technologies, and infrastructure for data generation, collection, modeling, and analysis;
4. Data cyberinfrastructure: Foster the creation of a robust, trustworthy, and performant data cyberinfrastructure and services that can support data-driven research and discovery in multiple areas of science and engineering; and
5. Education and workforce development: Develop coordinated activities in data science education, researcher training, and knowledge transfer, and harness the power of data at the local, state, national, and international levels in the service of science and society.

Each of these goals is designed to amplify the intrinsically multidisciplinary nature of the emerging field of data science. The HDR Big Idea will establish theoretical, technical, and ethical frameworks that will be applied to tackle data-intensive problems in science and engineering, contributing to data-driven decision making that impacts society.

FY 2020 Stewardship Investments

Foundations of Data Science ($6.0 million)
HDR will continue to support research in data science and data-enabled science and engineering primarily through the Transdisciplinary Research In Principles Of Data Science (HDR TRIPODS) program. HDR TRIPODS will bring together the electrical engineering, mathematics, statistics, and theoretical computer science communities. Through integrated research and training activities, these communities will collaborate to develop the theoretical foundations of data science. Phase I HDR TRIPODS awards will support the development of small, collaborative "data science institutes." Subsequent Phase II awards will enable the most successful of these smaller institutes to expand in scope and impact into a smaller number of larger-sized data science institutes.

Data-Intensive Research in Science and Engineering ($21.0 million)
HDR will continue to support the Institutes for Data-Intensive Research in Science and Engineering (DIRSE). The DIRSE institutes will complement the HDR TRIPODS institutes described above, and will harness diverse data sources and develop new algorithms, methodologies, systems, technologies, and infrastructure for data management and analysis to address critical national problems. These institutes will be developed through a two-phase process involving conceptualization followed by convergence. By creating a portfolio of interrelated institutes, NSF aims to accelerate discovery and innovation in multiple areas of data-intensive science and engineering.

Education and Workforce Development ($3.0 million)
HDR will continue to support data science education and workforce development through its Data Science Corps program. This program will provide data scientists and data science students with practical experiences, new skills, and teaching opportunities across multiple learning environments. The program will also strive to promote data literacy and provide basic training in data science to the existing workforce across various communities.

Additional FY 2020 Investments

HDR Track within the NSF Convergence Accelerator (NSF C-Accel)
The NSF C-Accel seeks to transform how the agency supports the most innovative science, reflecting its commitment to be at the cutting-edge, supporting foundational research, while also encouraging rapid advances through partnerships between academic and non-academic stakeholders. Tracks within the NSF C-Accel focus on grand challenge themes that would benefit from acceleration. To fully exploit the progression from data to knowledge to information, the HDR track within the NSF C-Accel will support research and development of an Open Knowledge Network (OKN). The OKN will allow stored data (both structured and unstructured data) to be located, and their attributes and relationships to other data and to real-world objects and concepts to be understood at a semantic level. A focus of the OKN will be on exploiting publicly available datasets from the U.S. Government and other sources. For more information on the NSF C-Accel, please refer to the NSF C-Accel narrative in this chapter.

Foundational Activities
These activities comprise ongoing investments by NSF directorates and offices in programs that laid the initial foundations for the HDR Big Idea and HDR Track in the C-Accel. These activities will continue to be supported and aligned with the overall HDR strategic goals. These foundational programs are currently managed by NSF's directorates and offices and will continue to remain within the directorates and offices with respect to their funding and management.