

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

HDR Funding (Dollars in Millions)			
	FY 2019 Actual	FY 2020 (TBD)	FY 2021 Request
Stewardship Activities (CISE)	\$30.01	-	\$45.00
Foundational Activities	\$144.39	-	\$112.11
BIO	7.80	-	7.41
CISE	64.60	-	52.49
EHR	5.15	-	2.50
ENG	21.87	-	11.97
GEO	5.60	-	2.85
MPS	32.16	-	19.10
SBE	7.21	-	5.79
IA	-	-	10.00
Total	\$174.40	-	\$157.11

Overview

NSF's HDR Big Idea is a national-scale activity to enable novel modes of data-driven discovery that will allow new fundamental questions to be asked and answered at the frontiers of science and engineering. HDR will generate new knowledge and understanding, and accelerate innovation. 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, simulation, and finally discovery and decision making, 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 the data revolution; and development of a 21st-century data-capable workforce. HDR will enable mutually beneficial interactions between data scientists and communities—supporting transfer of data science techniques to local communities, while providing insights and practical experience to participating data scientists and data science students in real-world settings. Importantly, HDR investments in FY 2021 will allow NSF to fund critical new methods and advances in artificial intelligence (AI), notably in machine/deep learning. The development and application of AI methods will further accelerate data-driven discovery in all fields of science, engineering, and education.

Goals

The HDR vision is realized through a set of interrelated 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 robust, trustworthy, and performant data cyberinfrastructure and services that can support data-driven research and discovery in multiple areas of science and engineering.
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.

FY 2021 Investments

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. In FY 2019, Phase I HDR TRIPODS awards (spanning 15 projects) were made to support the development of small, collaborative “data science institutes”. Subsequent Phase II awards, planned for FY 2022, will enable a subset of the most successful of these smaller institutes to expand in scope and impact into larger-sized data science institutes.

Data-Intensive Research in Science and Engineering (\$33.0 million)

HDR will support 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 science and engineering problems. In FY 2021, investments in HDR DIRSE institutes will emphasize fundamental research and education in AI, focused on machine/deep learning, that will accelerate data-driven discovery in all fields of science and engineering. The DIRSE institutes program constitutes a two-phase process involving conceptualization followed by convergence. In FY 2019, NSF issued more than 100 conceptualization awards spanning 28 projects supporting interdisciplinary teams that are conceptualizing and piloting new modalities for collaboration and convergence that go beyond traditional disciplinary and organizational boundaries. These projects will pave the way to the DIRSE convergence institutes planned for FY 2021. Furthermore, a coordination hub will be established in FY 2021 to support interaction, coordination, and sharing across the DIRSE institutes. By creating a portfolio of interrelated DIRSE institutes, NSF aims to accelerate discovery and innovation in multiple areas of data-intensive science and engineering.

Education and Workforce Development (\$6.0 million)

HDR will continue to support data science education and workforce development through the Data Science Corps (DSC) program. NSF funded 22 DSC awards spanning nine projects in FY 2019. These awards are helping to build the data science workforce by engaging data science students and professionals in real-world data science projects that will help bridge the data-to-knowledge gap in organizations and communities at local, state, national, and international levels. In FY 2021, the DSC program will continue to provide data science students and professionals with practical experiences, new skills, and teaching opportunities across multiple learning environments; promote data literacy, including its ethical usage; and provide basic training in data science to the existing workforce across communities throughout the United States. Furthermore, in FY 2021, investments in HDR DSC will enable education and workforce development of next-generation data scientists proficient in state-of-the-art AI systems and techniques.

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 Convergence Accelerator (CA). These activities will continue to be supported and aligned with the overall HDR strategic goals. These foundational activities 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.

HDR engagement with the Convergence Accelerator (CA)

In FY 2021, although not presented in the table above, the CA seeks to transform how the agency supports the most innovative science and engineering, 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 CA focus on grand challenge themes that would benefit from acceleration. HDR will continue its cooperation with the CA by building on the projects funded in FY 2019 as part of the HDR Open Knowledge Network track as well as through collaboration on awards funded through new tracks. For more information, refer to the CA narrative in this chapter.