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

HDR Funding[1]
(Dollars in Millions)

	FY 2020 Actual	FY 2021 Estimate	FY 2022 Request
Stewardship Activities (CISE)	**$30.00**	**$30.00**	**$30.00**
Foundational Activities	**$183.64**	**$143.86**	**$149.61**
BIO	10.00	7.41	7.41
CISE	59.60	59.60	59.60
EHR	17.40	2.50	2.50
ENG	13.90	12.75	13.50
GEO	5.60	5.60	5.60
MPS	26.89	22.00	22.00
SBE	9.76	7.25	7.25
TIP[2]	18.02	16.75	21.75
IA	22.47	10.00	10.00
Total	**$213.64**	**$173.86**	**$179.61**

[1] Funding dissplayed may have overlap with other topics and programs.

[2] FY 2020 and FY 2021 funding for TIP is shown for comparability across fiscal years.

Overview

HDR enables novel modes of data-driven discovery that will allow new fundamental questions at the frontiers of science and engineering to be asked and answered. It supports fundamental research in data science and engineering; development of a cohesive research data infrastructure needed to power the data revolution; and development of a diverse 21st-century data-skilled workforce. HDR will enable mutually beneficial interactions between data scientists and communities, thus 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.

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 to prepare a diverse workforce to harness the power of data at the local, state, national, and international levels in the service of science and society.

FY 2022 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 brings 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 were made to 15 projects, supporting the development of small, collaborative "data science institutes." In FY 2022, Phase II awards 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 ($21.0 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 2019, NSF issued more than 100 conceptualization awards spanning 28 projects that supported interdisciplinary teams to conceptualize and pilot new modalities for collaboration and convergence that go beyond traditional disciplinary and organizational boundaries. These projects paved the way for the DIRSE convergence institutes solicitation in FY 2021, which will fund DIRSE institutes in FY 2021 and their continued operation in FY 2022. 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 ($3.0 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. The DSC program continues to provide data science students and professionals with practical experiences, new skills, and teaching opportunities across multiple learning environments; promote data literacy, including the ethical use of data; and provide basic training in data science to the existing workforce across communities throughout the United States. A second round of DSC awards is planned for FY 2021 and these awards will continue in FY 2022. A focal point of the education and workforce development portfolio within HDR is to enable the participation of diverse backgrounds and perspectives in the future workforce.

Foundational Activities

These activities comprise ongoing investments by NSF directorates and offices in programs that laid the initial foundations for the HDR Big Idea. 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.