R&D LOGIC PRELIMINARY REPORT: ASSESSMENT OF FEDERAL R&D FUNDING STRATEGIES

Andrew Moore, Susan Chavez, Isaac Maya, Randolph Hall

November 30, 2023

Science and Technology Analysis and
Management of Innovation Activity (STAMINA III)

DHS Science and Technology (S&T)
Office of Science and Engineering (OSE)
Technology Scouting and Transition Division (TST)

Basic Ordering Agreement (BOA):
HSHQDC-17-A-B0004/70RSAT20FR0000097

Center for Risk and Economic Analysis of Threats and Emergencies (CREATE)
University of Southern California (USC)
1150 S. Olive St., 17th Floor
Los Angeles, CA 90015

This research was supported by the United States Department of Homeland Security (DHS) through the Center for Risk and Economic Analysis of Terrorism Events (CREATE) at the University of Southern California (USC) under Basic Ordering Agreement (BOA), HSHQDC-17-AB0004/70RSAT20 FR0000097. However, any opinions, findings, and conclusions or recommendations in this document are those of the authors and do not necessarily reflect views of the United States Department of Homeland Security, or the University of Southern California, or CREATE.

Center for Risk and Economic Analysis of Threats and Emergencies (CREATE)
The Nation’s First Homeland Security Center of Excellence

Table of Contents

1. EXECUTIVE SUMMARY

Federal agencies employ varying models for selecting and funding research and development (R&D) at universities, companies and other external performers. Utilizing public data sources, we first compare funding models as they relate to innovation. We second analyze metrics of innovation at universities as they relate to the amount and distribution of funding that they receive, by source. No confidential information or data is utilized or provided in this report.

2. R&D FUNDING MODELS AT FEDERAL AGENCIES

2.1 SECTION SUMMARY

In this section, we compare R&D funding models utilized by five federal agencies – DARPA, ARPA-H, DHS S&T, NASA and NSF – as they relate to innovation. Our comparison is based on public information, as described on agency websites. The unique features are summarized below.

DARPA (Defense Advanced Research Projects Agency):

- **Defense Focus**: DARPA primarily focuses on research and development efforts related to defense technologies, military applications, and addressing national security challenges.
- **High-Risk, High-Reward**: DARPA is known for funding high-risk, high-reward projects that have the potential for transformative impact but also carry significant technical challenges.
- **Rapid Innovation**: DARPA emphasizes quick development cycles, promoting agile and efficient technology advancements.
- **Short-Term Horizon**: DARPA often focuses on projects with relatively short-term objectives and aims to transition successful technologies to other entities for further development and deployment.

ARPA-H (Advanced Research Projects Agency for Health):

- **Proposed Agency:** ARPA-H is a proposed agency within the National Institutes of Health (NIH) that aims to accelerate breakthroughs in health research and innovation.
- **Health Focus**: ARPA-H would primarily focus on advancing health-related research, technology, and medical innovations to address critical health challenges and improve public health outcomes.
- **Long-Term Health Impact**: ARPA-H would likely pursue research projects with longer-term objectives and a focus on understanding and addressing complex health issues.
- **Integration of Multiple Disciplines:** ARPA-H would likely integrate expertise from various scientific fields, technology sectors, and healthcare domains to drive innovation in health research.

DHS S&T (Department of Homeland Security Science and Technology Directorate):

- **Homeland Security Focus**: DHS S&T primarily focuses on research and development efforts related to enhancing national security, disaster response, border protection, cybersecurity, and other areas related to homeland security.
- **Operational Applications:** DHS S&T aims to develop practical and deployable solutions that can be implemented by DHS and its operational components.
- **Collaboration with Stakeholders**: DHS S&T collaborates with industry, academia, and other government agencies to address the technological needs of the homeland security enterprise.

NASA (National Aeronautics and Space Administration):

- **Space and Aeronautics Focus:** NASA primarily focuses on space exploration, satellite missions, space science, aeronautics research, and technology development related to space and aviation.
- **Space Missions**: NASA leads and participates in missions to study Earth, the solar system, and the universe, including human spaceflight endeavors.
- **Collaboration and International Partnerships**: NASA collaborates with international space agencies, industry partners, and academic institutions to achieve its goals, including projects like the International Space Station (ISS).

National Science Foundation (NSF):

- **Broad Scientific Focus**: NSF supports research in various scientific disciplines, including mathematics, computer science, engineering, social sciences, natural sciences, and more.
- **Fundamental Research**: NSF focuses on supporting fundamental research that contributes to fundamental knowledge and understanding in scientific fields.
- **Merit-Based Funding**: NSF funds research projects based on competitive merit, peer review, and the potential for significant

2.2 BACKGROUND

We compared innovation related funding programs at five prominent federal R&D agencies: *Defense Advanced Research Projects Agency* (DARPA), *Department of Homeland Security Science & Technology* (DHS S&T), *National Science Foundation* (NSF), *National*

Aeronautics and Space Administration (NASA), and ARPA - H, with a focus on their respective approaches to research, projects, partnerships, common issues, and key performance indicators.

DARPA, the Defense Advanced Research Projects Agency, is renowned for its role in advancing cutting-edge technologies for national security purposes. ARPA-H is a proposed agency within the National Institutes of Health (NIH) that aims to accelerate breakthroughs in health research and innovation. The Department of Homeland Security Science and Technology Directorate (DHS S&T) prioritizes research and development efforts to address homeland security challenges. NASA pioneers space exploration and scientific discovery, while the National Science Foundation (NSF) supports fundamental research across a wide range of scientific disciplines.

We analyze various aspects of the selected agencies, including their mission, research activities, ongoing projects, employee profiles, approval processes for programs, organizational units, budget allocation, partnerships, common challenges, a notable project example, and key performance indicators (KPIs). By examining these aspects, we aim to understand how these agencies contribute to research and innovation within their respective domains, identify common challenges they face, highlight successful project examples, and explore the KPIs used to measure their performance. Through this analysis, we provide insights into the strategies, achievements, and unique characteristics of DARPA, ARPA-H, DHS S&T, NASA, and NSF shedding light on their contributions to research and innovation and the ways in which they shape the future of their respective fields.

2.3 RESULTS

Table 1 provides summary information on the five R&D agencies. In summary, DARPA, ARPA-H, DHS S&T, NASA, and the National Science Foundation (NSF) represent the diverse landscape of government agencies dedicated to research and innovation in the United States. Each agency has its own distinct mission, focus, and approach, contributing to advancements in various fields. NSF and ARPA-H are not operating agencies, responsible for internally implementing the innovations arising from their R&D support. Instead, their approach aims toward transfer of innovations into outside entities. DARPA, DHS and NASA have dual missions, both aiming for adoption within operating agencies (defense, homeland security or space) and also transfer into commercial entities.

DARPA stands out for its commitment to advancing cutting-edge technologies for national security purposes, while ARPA-H aims to accelerate breakthroughs in health research and innovation. DHS S&T focuses on developing practical solutions for homeland security challenges, NASA pioneers space exploration and scientific discovery, and the National Science Foundation supports fundamental research across a wide range of scientific disciplines. These agencies face various challenges, such as budget constraints, technological hurdles, coordination among stakeholders, and addressing complex scientific and societal issues. However, their commitment to overcoming these challenges and driving innovation remains unwavering.

Notable project examples show the impact these agencies have made in their respective domains. These projects demonstrate their ability to push scientific boundaries, develop transformative technologies, address critical societal challenges, and contribute to national security, health, space exploration, and fundamental scientific knowledge.

Key performance indicators play a vital role in assessing the effectiveness and progress of these agencies. These metrics enable the evaluation of project outcomes, collaboration effectiveness, budget utilization, and overall mission success, ensuring accountability and guiding future strategic decisions.

In conclusion, DARPA, ARPA-H, DHS S&T, NASA, and NSF each play a crucial role in advancing research and innovation within their domains. Their unique missions, approaches, partnerships, and contributions to scientific knowledge and societal impact shape the future of research and innovation in the United States. By fostering a culture of exploration, discovery, and progress, these agencies drive technological advancements, address complex challenges, and lay the foundation for a brighter future.

Table 1. Comparison of Federal R&D Agencies

	DHS S&T	ARPA - H	DARPA	NASA	NSF
Website Link	https://www.dhs.gov/science-and-technology	https://arpa-h.gov	https://www.darpa.mil	https://www.nasa.gov	https://www.nsf.gov
About	Fosters the tools, technologies, and knowledge products needed to secure the nation. They're in the business of identifying and developing innovative solutions.	Advances high-potential biomedical and health research that cannot be readily accomplished through traditional research or commercial activity.	Be the initiator and not the victim of strategic technological surprises	America's civil space program and the global leader in space exploration. **Vision:** Exploring the secrets of the universe for the benefit of all.	Independent federal agency that supports science and engineering in all 50 states and U.S. territories.
Mission	S&T's R&D is driven by the DHS's **Core missions includes:** - Counter terrorism and homeland threats - Secure U.S. borders - Secure cyberspace and critical infrastructure - Preserve and upload the nation's prosperity and economic security - Strengthen preparedness and resilience - Champion the DHS workforce and strengthen the department	ARPA-H accelerates better health outcomes for everyone by supporting the development of high-impact solutions to society's most challenging health problems.	Make pivotal investment in breakthrough technologies for national security through **transformational change** instead of incremental changes	NASA explores the unknown in air and space, innovates for the benefit of humanity, and inspires the world through discovery. **Key:** NASA focuses on technology programs more than the missions (as in the projects they do)	NSF was established in 1950 by Congress to: **Promote** the progress of science, **Advance** the national health, prosperity and welfare and **Secure** the national defense. NSF fulfills the mission by making grants. Their investments account for about 25% of federal support to America's colleges and universities for basic research. They also support solutions oriented research.

Research	Focuses on 3 types of solutions: **Near-term component:** projects that focus on gaps/needs that have been identified by DHS components **Foundational Science:** Enduring research that results in better and actionable data sets, knowledge products, standards, and peer-reviewed publications to support scientific endeavors. **Future Needs and Emerging Threats:** Exploring emerging science and technology areas and their potential threat or application to future DHS missions. **Purpose:** S&T delivers scientific and technical expertise to optimize decision-making, enable technical development, and enhance acquisition effectiveness for the homeland security enterprise.	ARPA-H provides research funding to build high-payoff capabilities or platforms to drive biomedical breakthroughs – ranging from the molecular to societal – that will provide transformative solutions for all individuals. The focus areas below illustrate the types of work and impact that ARPA-H may pursue as it hires its first PMs.	Investment strategy begins with a portfolio approach. Programs are infinite duration while creating revolutionary change **Purpose:** address a wide range of technology opportunities and national security challenges	Conduct research, test, and development to advance aeronautics, including electric propulsion and supersonic flight. Develop and fund space technologies that will enable future exploration and benefit life on Earth. The agency also shares what it learns so that its information can make life better for people worldwide. For example, companies use NASA discoveries and technologies to create new products for the public.	**Focus:** science & technology **Purpose:** becoming global leader in research and innovation
Example Projects	Aviation security to chemical and biological detection to critical infrastructure, resilience, climate and natural disasters, cybersecurity, and beyond	Examples of transformative progress - Cancer and Other Chronic Diseases - Healthcare Access, Equity and Quality	Game-changing military capabilities: precision weapons, modern civilian technology: internet, automated voice recognition, language translation	Technologies they're working on - Dual use technology - Helps develop and test new aircraft List for current projects: click here	NSF supports nearly two dozen large facilities, which are designed to serve the national and international science community at a scale requiring major investments

Employees/ Expertise	The program managers, scientists, engineers, technology specialists, and subject matter experts work directly with DHS component agencies, first responders at all levels, emergency management and public safety personnel, and operators in the field to understand their unique needs and challenges. They collaborate with partners from a broad network of federal, state, local, tribal and territorial governments; national laboratories; industry innovators; academia and international agencies to pinpoint capability gaps and build technologies and publish guidance.	*Couldn't find information * ARPA-H is a new agency within the National Institutes of Health (NIH)	220 government employees in 6 technical offices, ~100 program managers who together oversee ~250 research and development programs **Program managers**: top of fields: academia, industry and gov/ agencies for limited stints, generally 3 to 5 years. Address challenges broadly spanning from deep science to systems to capabilities - Define programs, set milestones, meet w/ performers & track progress **Report:** DARPA's office directors and deputies	Under 18,000 civil servants, and works with many more U.S. contractors, academia, and international and commercial partners to explore, discover, and expand knowledge for the benefit of humanity. NASA supports more than 312,000 jobs across the United States, generating more than $64.3 billion in total economic output (Fiscal Year 2019).	Each research area is headed by an assistant director and each is further subdivided into divisions like materials research, ocean sciences and behavioral and cognitive sciences. Within NSF's Office of the Director, the Office of Integrative Activities also supports research and researchers. Other sections of NSF are devoted to financial management, award processing and monitoring, legal affairs, outreach and other functions.
Approval of programs	S&T's work in systems engineering and standards, test and evaluation, and operational analysis has made S&T a trusted resource.	Partnerships are directly with the government	DARPA Director and Deputy Director approve each new program and review on going programs, while setting Agency-wide priorities and ensuring a well balanced investment portfolio	Many NASA partnerships are attributable to direct communication between the potential partner and a NASA Center and are not derived from a formal Partnership Announcement	Partnerships are formalized through Memoranda of Understanding (MOUs) between NSF and the partner(s).

Units / Directories / Projects	**Topics of Research** - Border Security - Chem, Bio & Explosive Defense - Counter Terrorist - Cybersecurity / Information Analysis - First Responder / Community and infrastructure Resilience - Food and Agriculture Defense - Physical Security and Critical Infrastructure Resilience - S&T Directorate's COVID-19 Reponse	**Focus Areas** - Health Science Futures - Scalable Solutions - Proactive Health - Resilient Systems	Research portfolio is managed by 6 technical offices that manage the special projects and the transition of DARPA-funded technologies into Department of Defense Capabilities - Biological Technologies Office - Defense Sciences Office - Information Innovation Office - Microsystems Technology Office - Strategic Technology Office - Tactical Technology Office	**7 research topics** - Humans in Space - Moon to Mars - Earth - Space Tech - Flight - Solar System and Beyond - STEM Engagement	NSF is divided into seven directorates that support science and engineering research and education: - Biological Sciences - Computer and Information Science and Engineering - Engineering - Geosciences - Mathematical and Physical Sciences, - Social, Behavioral and Economic Sciences - STEM Education - Technology, Innovation and Partnerships - Human Resources - Integrative Activities - International Science and Engineering (ISE)
Budget	**FY2023:** $901.3M **FY2022:** $822.9M **FY2021:** $765.6M	**FY2024:** $2.5 billion **FY2023:** $1.5 billion **FY2022:** $1 billion	**FY2023:** $4.119 billion **FY2022**: $3.868 billion	**FY2023:** ~$25.3 billion **FY2022:** ~$24 billion **FY2021:** $23.2 billion **Science & Technology** **FY2023:** ~$3.6 billion	**FY2024**: NSF's budget request to Congress is $11.3.14 billion (18.6% from agency's current budget) **FY2023**: Request of $10.492 billion **FY2022**: $8.8 billion (a 4.1% appropriation) increase above its FY 2021

Partnership opportunities / Criteria for funding	**Partnership Guide** Relies on private sector partners to help discover scientific advancements and technological innovations that solve homeland security challenges. There are many opportunities to work with S&T, from a wide variety of open solicitations such as the Long Range Broad Agency Announcement to specific programs like the SAFETY Act that provide incentives for the development and deployment of anti-terrorism technologies. **Engage** S&T's Industry Liaison is your primary entry point into S&T (SandT.Innovation@hq.dhs.gov.) **Develop** S&T's innovation Funding Program and Rules are unique for working with many kinds of entities **Deliver** S&T's Technology Transfer and Commercialization (T2C) is the central point to manage technology transfer activities throughout DHS and the DHS labora **More information on pages 17 & 18 of Partnership Guide**	At present minimal information is available beyond this: **Opportunity**	**Opportunities** - Academic - Corporate - Gov. partners DARPA publicizes funding opportunities by posting Broad Agency Announcements (BAAs) that formally request proposals tied to specific areas of research and development. R&D is focused mainly on the 6 offices, but they could consider areas outside of those offices **DARPA Opportunities:** DARPA BAAs, Special Notices, Requests for Proposals, and Requests for Information can be found on the federal acquisition opportunities website, SAM.gov. Some DARPA opportunities may also be found at https://www.grants.gov/. For DARPA Small Business Innovation Research (SBIR)/Small Business Technology Transfer (STTR) opportunities, please see For Small Business. DARPA's Young Faculty Award program offers funding to promising junior faculty members and their peers at nonprofit research institutions.	**NASA Partnerships** Explore Partnerships Steps to achieving a NASA Partnership NASA offers its unique capabilities and resources for use by - Commercial industries - Academic institutions - U.S. Government agencies - International entities NASA has different partners for their projects Science and Technology Funding Sources: Link	**Landscape Study** Partnerships are formalized through Memoranda of Understanding (MOUs) between NSF and the partner(s). NSF stimulates partnerships indirectly through its many programs that require or encourage grantees to work in collaboration with non-academic entities **Pg. 2:** More details about how NSF's Directorates and Offices choose partnerships **Direct Partnerships** - Interagency (U.S. government) - Industrial - Private foundation and NGO - International **NSF-Catalyzed Partnerships** - Many NSF programs expect grantees to work with non-academic partners Many of NSF's direct partnerships are formalized through Memoranda of Understanding (MOUs) or other signed agreements, while others are informal.

Partnerships	- Federally Funded Research and Development Centers - Industry Partnerships - International Partnerships - University Programs - National Laboratories	- Government	- Industry - Small businesses - Universities - Government & Military	- Strategic University Research Partnerships - Industry Partnerships *Note: some partnerships are determined per project*	- Federal agencies, industry, - Private foundations & NGOs - Foreign funders
University Partnership	Click on link for more information on how to apply for funding - Programs	The main partners are the "Government" but they consider "Industry and Academia" as the "Performers"	Click on link for more information on how to apply for funding - Universities	Click on link for more information on how to apply for funding - Strategic University Research Partnerships Science and Technology Partnerships - Link	Click on link for more information on how to apply for funding - Universities
Industry Partnership	Click on link for more information on "Industry Partnership" - Industry	No Industry Partners	Click on link for more information on "Industry Partnership" - Industry	Click on link for more information on "Industry Partnership" - Industry	Click on link for more information on "Industry Partnership" - Industry
Common issues	**Complex Threat Landscape:** new and emerging threats continuously arise so it can be difficult staying ahead of the evolving threats and developing effective countermeasures **Technology Development and Adoption:** It can be difficult to develop and deploy new technologies as they must navigate through bureaucratic processes and technology procurement procedure **Interagency Coordination:** collaborating with different partners can lead to	**Funding**: The projects are high-risk, high0reard research projects so ensuring a consistent and sufficient budget can be a challenge **Stakeholder Alignment**: Coordinating and aligning the interests and priorities of various stakeholders, including researchers, scientists, policymakers, and industry partners, can be complex. **Regulatory and Ethical Considerations**: Health research often involves ethical and regulatory complexities, such as ensuring patient privacy,	**Technological and Scientific Challenges**: DARPA's tackles cutting-edge technologies which can lead to technical challenges and uncertainties, as breakthroughs in these areas are not guaranteed **Rapid Technological Evolution**: Staying ahead of the curve and anticipating future needs requires constant monitoring of technological developments and proactive planning. **Coordination and Collaboration**: Coordinating partnerships, aligning objectives, and managing	**Budget Constraints**: Limited financial resources can impact NASA's ability to execute missions and research projects as planned. **Technical Complexity**: Developing and operating spacecraft, conducting experiments, and managing mission logistics require overcoming intricate engineering challenges and complex technologies. **Safety and Risk Management:** Ensuring the safety of astronauts and spacecraft in the face of inherent risks associated with space exploration is of paramount importance to	**Funding Constraints**: The challenge of securing sufficient funding to support the wide range of research and educational programs. **Shifting Research Priorities**: Adapting to evolving scientific landscapes and changing societal needs to align research priorities with emerging challenges. **Interdisciplinary Collaboration**: Overcoming barriers and fostering effective collaboration between researchers from different disciplines to address complex scientific and societal challenges. **Public Engagement and Science Communication**:

	difficulties with coordinating efforts, sharing information, aligning priorities **Technology Transition and Commercialization:** Successfully transitioning R&D efforts from laboratory to operation use can be challenging	informed consent, and adherence to ethical guidelines. **Technology Transfer and Commercialization**: Translating research findings and innovative technologies into practical applications and commercial products can be a significant challenge. **Public Perception and Engagement:** Building public trust and awareness around the work of ARPA-H is essential.	expectations can be complex, particularly when working on multidisciplinary projects that involve diverse organizations and interests. **Transitioning Research to Practical Applications**: The focus on cutting-edge research and technology can make it challenging to transition projects from the laboratory to practical applications in the field. The agency must navigate between successful research outcomes and the practical adoption of those outcomes by end-users.	NASA. **Political and Policy Shifts**: Changes in government priorities, policies, or administration can affect NASA's long-term plans, funding stability, and direction of space exploration.	Effectively communicating scientific concepts to diverse audiences and engaging the public in scientific endeavors to promote scientific literacy and support.

Example project	Automated Targeting System **Project:** ATS is an advanced data analysis and risk assessment tool used by U.S. Customs and Border Protection (CBP) to help identify potential security risks among individuals and cargo entering the United States. **Results:** The ATS project has been successful in enhancing border security by improving the accuracy and efficiency of risk assessment processes. **Potential KPI's:** risk identification accuracy, false positive rate, processing time, targeted screening effectiveness, threat detection rate, operational efficiency	*Couldn't find information about ongoing/completed projects* One example of a project on their website is about "Novel Innovations for Tissue Regeneration in Osteoarthritis" **Project**: Novel Innovations for Tissue Regeneration in Osteoarthritis (NITRO) aims to address current issues surrounding osteoarthritis treatment by developing new ways of helping the human body repair its own joints. In particular, the program focuses on three technical areas: injectable bone regeneratives, injectable cartilage regeneratives, and replacement joints built from human cells. **Results**: currently seeking partnerships to work on this project to ultimately reduce health costs, increase access, and drastically improve the lives of millions of people with or at risk for osteoarthritis.	Precision Urban Hopper (PUH) **Project**: The purpose of the Precision Urban Hopper (PUH) project was to develop a robotic system for agile and autonomous movement in urban environments, assisting military forces in reconnaissance, surveillance, and operations in complex urban settings. **Results**: The results of the PUH project demonstrated the successful development of a hopping robot prototype capable of agile navigation in urban environments, showcasing improved mobility and maneuverability for military operations in complex urban terrains. **Potential KPI's:** Mobility, Stability and Balance, Range and Endurance, Autonomy, Payload and Functionality, User Feedback, Technology Transfer Potential	Mars Perseverance Rover mission **Purpose:** search for signs of past life, study the Martian geology and climate, and lay the groundwork for future human exploration of Mars. **Results**: The Mars Perseverance Rover mission is currently analyzing data to uncover valuable insights into Mars' geology and potential for past life. The ongoing mission continues to gather data to further our understanding of Mars' past and potential for supporting life. **Potential KPI's**: Financial performance, Customer satisfaction, Productivity, Quality, Time, Innovation, Employee engagement, Safety, Environmental sustainability, Market share	AI Institute for Artificial Intelligence and Fundamental Interactions **Results**: Aims to advance the field of AI and deepen the understanding of the fundamental interactions of the universe, with potential results including the development of novel AI techniques **Purpose**: Bring together researchers from AI and particle physics to explore the synergies between these fields, fostering collaboration, and leveraging AI technologies to accelerate scientific discoveries and advancements. **Potential KPIs**: number of interdisciplinary research papers published, successful collaborations between AI and particle physics researchers, contributions to AI algorithms and models

Key performance indicators	DHS has a structures process that aligns with the goals and objectives of the agency 1. Identify objectives 2. Determine Critical Success Factors 3. Identify KPIs 4. Establish Baselines and Targets 5. Define Data Collection and Reporting 6. Implement Monitoring and Evaluation 7. Continuous Improvement	*Couldn't find information about the org specific structure but will provide a general overview* 1. Clarify objectives 2. Align with strategic priorities 3. Consult stakeholders 4. SMART criteria 5. Data availability 6. Balanced metrics 7. Regular review and adaptation	General structure 1. Defining Objectives 2. Identifying Metrics 3. Stakeholder Engagement 4. Technical Assessments 5. Iterative Refinement 6. Program Review	Hierarchical structure 1. Strategic goal alignment 2. Stakeholder engagement 3. Data analysis 4. Performance metric selection 5. KPI definition 6. Review and refinement 7. Monitoring and evaluation	Structured approach 1. Goal Identification 2. Metrics Selection 3. Data Collection 4. Data Analysis 5. Performance Assessment 6. Review and Refinement 7. Reporting and Communication

3. INNOVATION INDICATORS OF SUCCESS

3.1 SECTION SUMMARY

We investigated the relationship between R&D spending at universities and metrics of university innovation. Toward that end, we integrated data from the National Science Foundation HERD (Higher Education Research and Development) dataset tracking R&D expenditures by year, institution and source (including federal and non-federal), with the AUTM (Association of University Technology Managers) STATT database, tracking licenses, royalty income, invention disclosures and patents by institution. The novel integrated dataset enabled analysis of relationships between investment and innovation metrics over time through multivariate statistical analysis. Our focus was on understanding the relationship between R&D expenditures to universities by funding source and two innovation metrics: licensing income and licenses issued.

Both HERD and STATT contain data annually self-reported by universities following specified templates. Neither dataset contains information on individual projects. Instead, they provide aggregated data among all projects in a given category. HERD data are collected by the National Science Foundation, which openly publishes data by individual university roughly 1.5 years after the end of each fiscal year (ending June 30). STATT data are collected by a university association and made available to contributing members, but are not openly published.

We note that neither HERD nor STATT contains data specifically identifying DHS S&T as a funding source. DHS is a component of "other federal funding" in HERD, whereas STATT does not delineate technology transfer metrics by funding sources.

3.2 METHODS

Data Preparation:

We combined the HERD data on university spending on a federal level and STATT database's metrics on university innovation. One of the major issues in combining the datasets is how each individual database records the name of the university. For example, the University of Southern California was identified as "U. of Southern California" in one database and "University of Southern California" in another. A second major issue was the level of aggregation within public university systems. In one dataset, data were reported at the system level whereas data were reported at the campus level in another. Naming conventions and system-level aggregation were not always consistent from year to year as well. A third major issue is that some universities did not consistently contribute to the STATT databased, resulting in missing data.

We used a few tools to assist in the data cleaning process:

- FuzzyWuzzy Data Matching AI Algorithm
- Regex
- Pandas

FuzzyWuzzy identifed rough matches of two columns within both databases. We also have implemented *Regex*, a natural language processing platform that makes it easier to correct syntax differences between the two databases. Finally, we implemented *Pandas*, which allows us to shape and reshape data frames based on the two databases. Furthermore, human interpretation was used to compare the two databases to validate matches. With these methods, we made the following changes.

HERD Database:

1. Combed all the Texas schools to make a University of Texas System
2. Combined all the California schools to make the University of California System
3. Corrected U. to University in the names of universities using natural language processing
4. Corrected C. to College in the names of colleges using natural language processing

STATT Database:

1. Removed entities that aren't in the HERD database
2. Corrected of university names to match HERD database

We combined these results for 2015-2020, representing six years of university based data.

<u>Statistical Analysis</u>

We utilized both multi-variable linear regression and correlation analysis to understand the relationships between innovation metrics (derived from STATT) and R&D expenditures (derived from HERD). Our aims were to understand whether it can be statistically shown that R&D expenditures produced more significant results from some sources than others. We also sought to identify which universities produced better than predicated outcomes, as a potential reflection of a more effective innovation strategy.

Our two primary indicators of success (IoS) from the STATT database were gross licensing income and total number of licenses. Gross Licensing Income is the total licensing income from license issue fees, payments under options, annual minimums, and running license income. The total number of licenses would be counting the number of licenses or options agreements that were executed in the year for that particular university. We used federal governmental agencies (DOD, HHS, NASA, and NSF), state and local governments, Business Funding, and institution funds as predictors for our two dependent variables. Through our analysis, we will be able to see what were the highest predictors for the indicators of success that we defined. We used the following metrics in the paper to define our results.

Metric	**Definition**	**Website**
R^2	Proportion of variance explained by the model (valued between 0 and 1).	Link
F-Statistic	A test statistic for model significance. The associated p-value provides the probability of observing such a value or larger if there had been no actual association for the given sample size.	Link
Z-Score	Tells you how far away a particular data point is from the average (mean) of a group of data points, measured in terms of standard deviation.	Link

Correlation Matrices

In addition to linear regression, we analyzed data for correlation between R&D expenditures by individual source and both gross licensing income and number of licenses executed, averaged among all years. We found that from the federal government, NSF and HHS had the highest correlations for both dependent variables. We also found from non-governmental funding that business funding had the highest correlation with gross licensing income, while institution funds had the highest correlation with total number of licenses. We also note that expenditures among sources show significant multi-collinearity, meaning that universities that receive large funding in one category tend to receive high funding in others. However, DOD and NASA funds tend not to correlate as highly with other sources, except with each other.

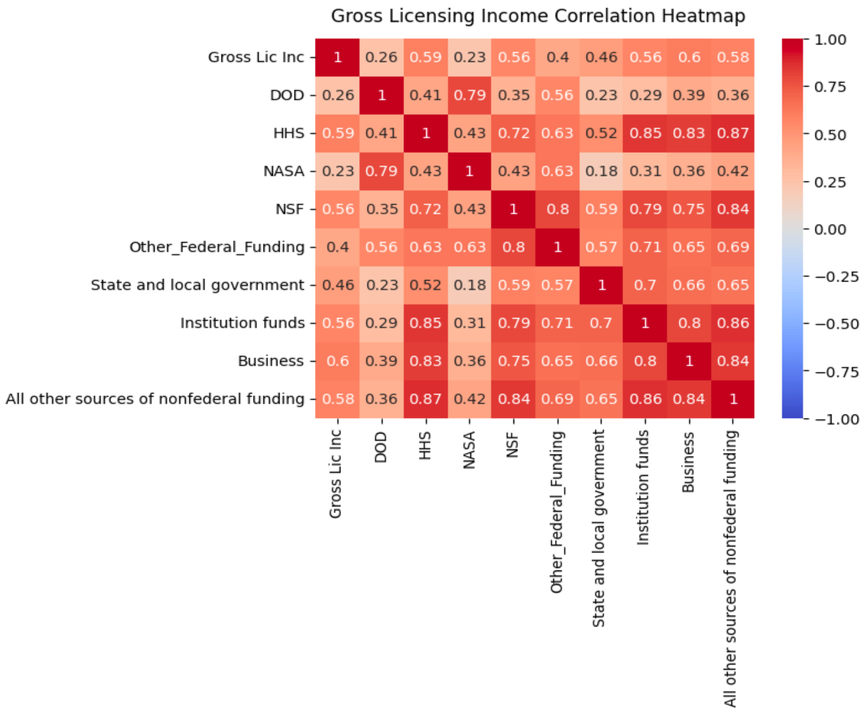

Figure 1: Gross Licensing Income Correlation Heat Map

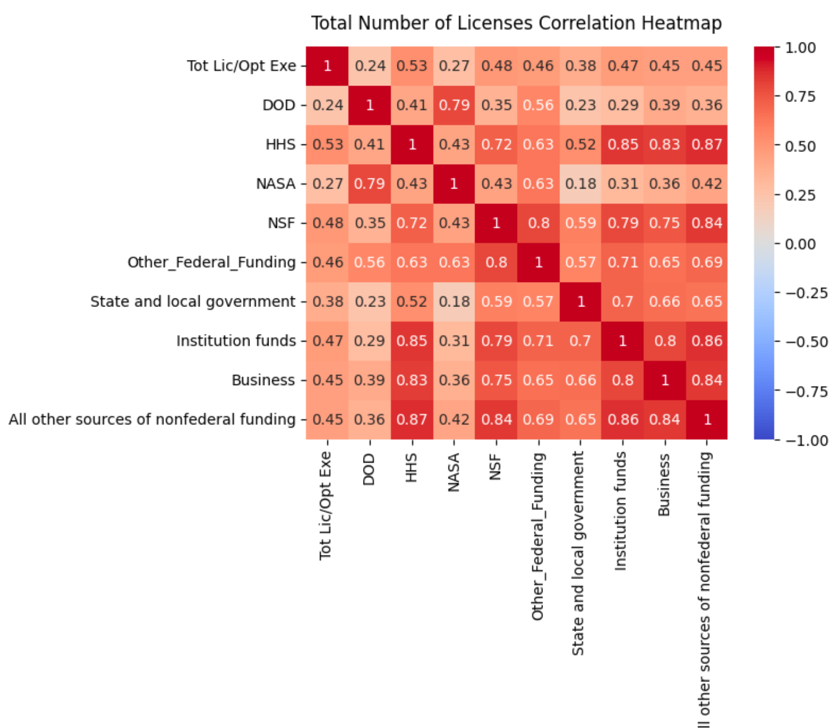

Figure 2: Total Number of Licenses Correlation Heat Map

We then looked at correlations between other IoS, like invention disclosures received (indiv), patent applications filed (nptapp), issued US patents (iss US pat), count of new startups (strup), total number of licenses (Tot Lic/Opt Exe), and gross licensing income (Gross Lic Income). We then compared these IoS with our independent variables. NSF funding generally showed the highest correlations with the IoS. The highest correlations were found with front-end technology transition activities, particularly invention disclosures, which are less dependent on market success. Lower correlations were found with the outcomes of executed licenses and patent income, which more closely reflect adoption of new technology.

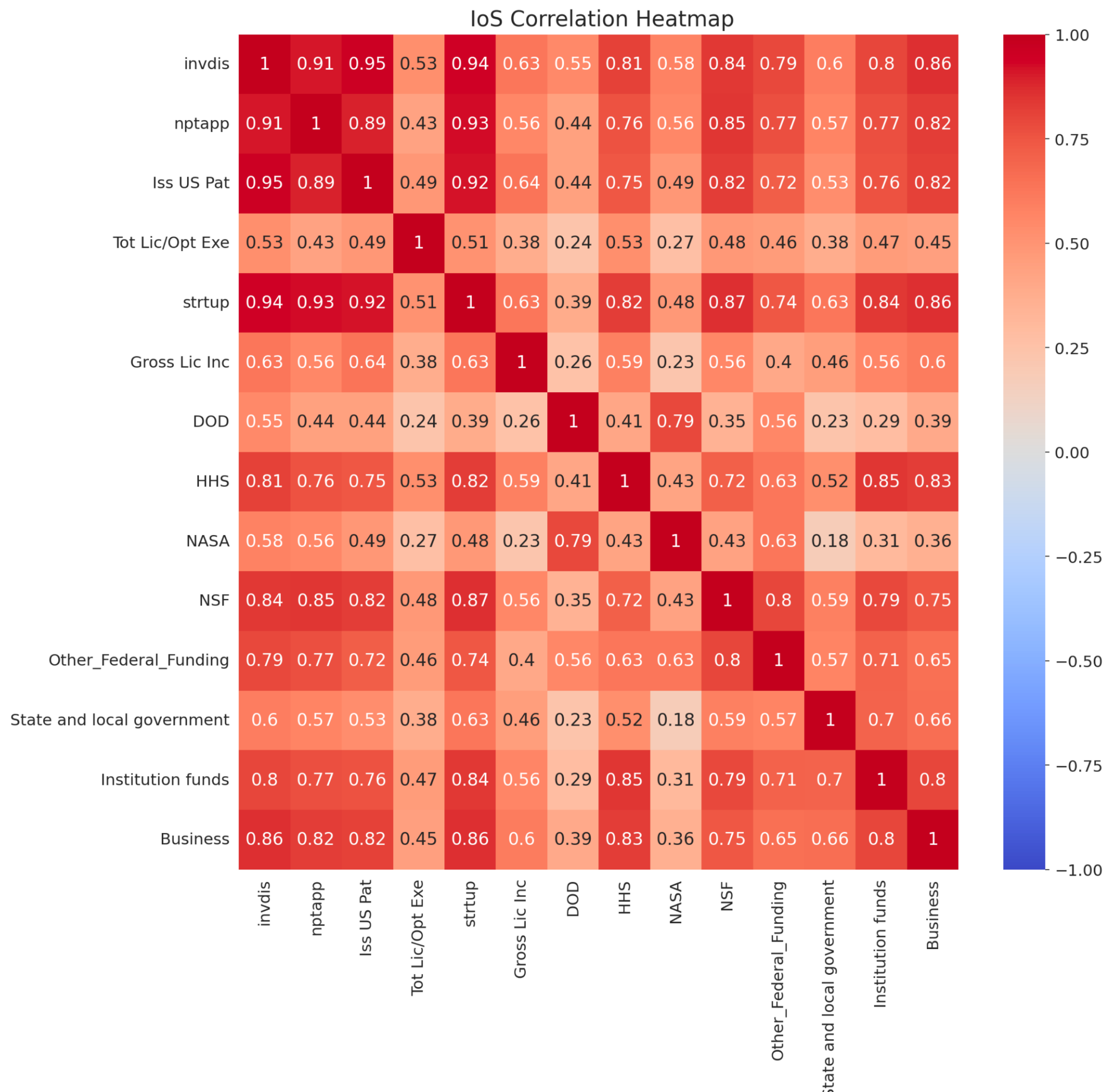

Figure 3: IoS Correlation Heat Map

3.3 REGRESSION RESULTS

Multivariate linear regression models were created for two dependent innovation variables: gross licensing income and total number of licenses. We averaged data for each university from 2015-2020 based on the number of years that the university appeared in both datasets. We also combined Department of Energy and United States Department of Agriculture data with Other Federal Funding, given these federal agencies worsened the fit of the model and provide relatively small funding compared to other sources. The results for gross licensing income and the total number of licenses follow in Tables 2 and 3.

Table 2: Gross Licensing Income as the Dependent Variable

OLS Regression Results

Dep. Variable:	Gross Lic Inc	R-squared:	0.437
Model:	OLS	Adj. R-squared:	0.410
Method:	Least Squares	F-statistic:	16.04
Date:	Fri, 26 May 2023	Prob (F-statistic):	2.67e-19
Time:	17:02:38	Log-Likelihood:	-3575.2
No. Observations:	196	AIC:	7170.
Df Residuals:	186	BIC:	7203.
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.103e+06	2.05e+06	0.538	0.591	-2.94e+06	5.15e+06
DOD	28.5107	28.513	1.000	0.319	-27.741	84.762
HHS	42.4352	19.633	2.161	0.032	3.703	81.168
NASA	-17.3992	110.743	-0.157	0.875	-235.873	201.075
NSF	286.3451	88.453	3.237	0.001	111.846	460.844
Other_Federal_Funding	-201.3117	78.056	-2.579	0.011	-355.301	-47.322
State and local government	82.4307	50.339	1.638	0.103	-16.878	181.739
Institution funds	2.6667	27.855	0.096	0.924	-52.286	57.619
Business	117.3717	67.465	1.740	0.084	-15.724	250.467
All other sources of nonfederal funding	-35.4455	43.441	-0.816	0.416	-121.146	50.255

Omnibus:	243.866	Durbin-Watson:	2.135
Prob(Omnibus):	0.000	Jarque-Bera (JB):	12390.605
Skew:	5.148	Prob(JB):	0.00
Kurtosis:	40.566	Cond. No.	4.19e+05

Table 3: Total Number of Licenses as the Dependent Variable

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:        Tot Lic/Opt Exe   R-squared:                       0.348
Model:                            OLS   Adj. R-squared:                  0.316
Method:                 Least Squares   F-statistic:                     11.01
Date:                Fri, 26 May 2023   Prob (F-statistic):           1.08e-13
Time:                        17:02:38   Log-Likelihood:                -1061.8
No. Observations:                 196   AIC:                             2144.
Df Residuals:                     186   BIC:                             2176.
Df Model:                           9                                         
Covariance Type:            nonrobust                                         
============================================================================================================
                                           coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------------------
const                                    6.2386      5.532      1.128      0.261      -4.675      17.152
DOD                                  -5.609e-05   7.69e-05     -0.729      0.467      -0.000    9.56e-05
HHS                                      0.0003    5.3e-05      4.736      0.000       0.000       0.000
NASA                                  4.708e-06      0.000      0.016      0.987      -0.001       0.001
NSF                                      0.0005      0.000      1.963      0.051   -2.25e-06       0.001
Other_Federal_Funding                    0.0003      0.000      1.465      0.145      -0.000       0.001
State and local government               0.0003      0.000      2.093      0.038    1.64e-05       0.001
Institution funds                       -0.0001   7.51e-05     -1.356      0.177      -0.000    4.63e-05
Business                             -9.664e-05      0.000     -0.531      0.596      -0.000       0.000
All other sources of nonfederal funding -0.0003      0.000     -2.348      0.020      -0.001   -4.39e-05
==============================================================================
Omnibus:                      311.092   Durbin-Watson:                   1.872
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            47410.600
Skew:                           7.409   Prob(JB):                         0.00
Kurtosis:                      77.738   Cond. No.                     4.19e+05
==============================================================================
```

We then looked at which institutions in the linear regressions had the highest ratios and smallest ratios of the actual values vs the predicted. Tables 4 and 5 show the top five highest performing institutions for gross licensing income and the total number of licenses.

Table 4. Highest Ratios for Gross Licensing Income

Rank	**University**	**Actual/Prediction Ratio**
1	Northwestern University	8.2
2	Carnegie Mellon University	7.2
3	Rockefeller University	6.8
4	Cold Spring Harbor Laboratory	6.5
5	New York University	5.3

Table 5: Highest Ratios for Total Number of Licenses

Rank	**University**	**Actual/Prediction Ratio**
1	University of Oregon	25.8
2	Ball State University	7.1
3	Massachusetts Institute of Technology	6.1
4	University of New Hampshire	4.9
5	University of Georgia	4.6

Northwestern University had the highest ratio in the gross licensing income, owing to a highly successful drug license. University of Oregon had the highest ratio for the total number of licenses. All 10 of the identified universities merit future investigation as to successful mechanisms for technology transition.

Modified Regression with Outliers Removed

We sought to improve model fit by excluding outliers using Z-scores. For each column, we calculated the Z-scores for every datapoint and excluded data points that were greater/less than +-2. This new model achieved higher R^2 values versus the original model, specifically in the total number of licensed models (R^2 of 0.348 versus 0.544). The results of this model follow in Table 6 and 7.

Table 6: Gross Licensing Income as the Dependent Variable, Outliers Removed

OLS Regression Results

Dep. Variable:	Gross Lic Inc	R-squared:	0.305
Model:	OLS	Adj. R-squared:	0.265
Method:	Least Squares	F-statistic:	7.611
Date:	Fri, 26 May 2023	Prob (F-statistic):	3.23e-09
Time:	17:02:43	Log-Likelihood:	-2864.2
No. Observations:	166	AIC:	5748.
Df Residuals:	156	BIC:	5780.
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-1.982e+04	8.98e+05	-0.022	0.982	-1.79e+06	1.75e+06
DOD	-7.9297	45.626	-0.174	0.862	-98.053	82.194
HHS	12.2719	14.140	0.868	0.387	-15.658	40.202
NASA	-170.3083	89.090	-1.912	0.058	-346.287	5.670
NSF	157.4703	66.850	2.356	0.020	25.423	289.517
Other_Federal_Funding	-47.2271	52.055	-0.907	0.366	-150.052	55.597
State and local government	11.6074	50.991	0.228	0.820	-89.114	112.329
Institution funds	20.0597	17.386	1.154	0.250	-14.284	54.403
Business	-11.3428	82.273	-0.138	0.891	-173.857	151.171
All other sources of nonfederal funding	89.2910	47.992	1.861	0.065	-5.508	184.090

Omnibus:	154.900	Durbin-Watson:	1.970
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2292.246
Skew:	3.520	Prob(JB):	0.00
Kurtosis:	19.788	Cond. No.	2.13e+05

Table 7: Total Number of Licenses as the Dependent Variables, Outliers Removed

OLS Regression Results

Dep. Variable:	Tot Lic/Opt Exe	R-squared:	0.544
Model:	OLS	Adj. R-squared:	0.518
Method:	Least Squares	F-statistic:	20.90
Date:	Fri, 26 May 2023	Prob (F-statistic):	6.34e-23
Time:	17:02:43	Log-Likelihood:	-737.50
No. Observations:	168	AIC:	1495.
Df Residuals:	158	BIC:	1526.
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.9441	2.359	-0.400	0.689	-5.602	3.714
DOD	-0.0002	0.000	-1.805	0.073	-0.000	1.94e-05
HHS	2.055e-05	3.51e-05	0.585	0.559	-4.88e-05	8.99e-05
NASA	0.0001	0.000	0.526	0.600	-0.000	0.001
NSF	0.0003	0.000	1.895	0.060	-1.23e-05	0.001
Other_Federal_Funding	0.0002	0.000	1.473	0.143	-6.7e-05	0.000
State and local government	0.0003	0.000	1.896	0.060	-1.06e-05	0.001
Institution funds	6.583e-05	4.1e-05	1.604	0.111	-1.52e-05	0.000
Business	0.0006	0.000	2.787	0.006	0.000	0.001
All other sources of nonfederal funding	5.118e-05	0.000	0.499	0.619	-0.000	0.000

Omnibus:	98.599	Durbin-Watson:	1.768
Prob(Omnibus):	0.000	Jarque-Bera (JB):	568.092
Skew:	2.177	Prob(JB):	4.37e-124
Kurtosis:	10.886	Cond. No.	2.26e+05

Given the improved score of both the R^2 value and the F-Statistic in the total number of licenses -- outliers removed – model, we can conclude that outliers play a significant role in prediction, as summarized below. By excluding outliers, the modified model provides a better fit for the remaining universities.

Dependent Variable and Type of model	**R^2 Value**	**F-Statistic**
Gross Licensing Income	0.437	16.04
Gross Licensing Income, Outliers Removed	0.348	11.01
Total Number of Licenses	0.305	7.611
Total Number of Licenses, Outliers Removed	0.544	20.90

3.4 DISCUSSION

Results show that NSF has the highest impact with the indicators of success (IoS), an interesting outcome given that among federal agencies NSF is the most focused on fundamental research. NSF differs in many more respects. It tends to fund in relatively modest increments; it is highly focused on engineering and science rather than medicine; it is not an operating agency, meaning technology transition is highly dependent on technology transfer, rather than internal adoption. Furthermore, NSF has a longstanding emphasis on broader impacts (benefiting society and contributing to desired societal outcomes), rewarding a variety of transitional outcomes.

3.5 CONCLUSIONS AND FUTURE STEPS

Our results showed that NSF funding had the most impact on our IoS. We also found that the gross licensing model had the highest R^2 value and F-Statistic. Furthermore, we found that outliers had a significant effect on the prediction of the number of licenses and removing these outliers produced a significantly better fit. The outliers are nevertheless important as they may represent effective technology transfer practices at individual universities.

In future work, we aim to combine more years of data for analysis for a multi-dimensional linear regression. We will extend the data cleaning pipeline for past years to capture additional syntax changes. Furthermore, due to the complexity of predicting the dependent variables, we will look into using tree-based methods for prediction. These tree-based methods including Random Forest Regressor and XGBoost to detect smaller discrepancies (and non-linearities) that might go unnoticed with simpler models. Finally, the robustness of tree-based methods to outliers makes them a very useful method for looking at our dataset.

Our future work will also combine analysis of agency R&D practices with indicators of technology transfer success.