CENTER FOR DATA INNOVATION

June 29, 2025

Dear Senators Heinrich and Rounds,

On behalf of the Center for Data Innovation (datainnovation.org), I am pleased to submit this response to the request for information from Senators Heinrich and Rounds on the development of the American Science Acceleration Project (ASAP), a bipartisan initiative focused on accelerating scientific progress by strengthening the structural, computational, and institutional foundations of U.S. research.[1]

The Center for Data Innovation studies the intersection of data, technology, and public policy. The Center formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. It educates policymakers and the public about the opportunities and challenges associated with data, as well as technology trends such as open data, artificial intelligence (AI), and the Internet of Things. The Center is part of the Information Technology and Innovation Foundation (ITIF), a nonprofit, nonpartisan think tank.

We thank the Senators for their leadership in advancing this agenda and welcome continued dialogue as the ASAP initiative develops.

Yours sincerely,

Hodan Omaar
Senior Policy Manager
ITIF's Center for Data Innovation

[1] Senator Martin Heinrich and Senator Mike Rounds, *American Science Acceleration Project: Request for Information*, June 2025. https://www.heinrich.senate.gov/newsroom/press-releases/heinrich-rounds-seeking-public-input-on-initiative-to-accelerate-advancements-in-american-science-asap.

EXECUTIVE SUMMARY

ASAP is a timely effort to modernize the infrastructure that powers U.S. research. By improving how science is organized, resourced, and executed, ASAP can help unlock faster breakthroughs across a range of disciplines. To support that mission, we offer six recommendations:

A. **Support Self-Driving Labs (SDLs):** To increase scientific productivity, particularly in materials science, Congress should support the deployment of automated research facilities that combine AI and robotics to better support scientific experimentation.

B. **Redesign scientific workflows around human insight:** Rather than focusing on keeping scientists at the center of the scientific workflow, Congress should prioritize the redesign of workflows to strategically offload core scientific tasks to machines, allowing human researchers to focus on designing inquiries, interpreting complex results, and generating new insights.

C. **Track AI adoption and workforce readiness:** Congress should direct the National Center for Science and Engineering Statistics (NCSES) to enhance its existing surveys to measure institutional uptake of AI tools and evaluate workforce proficiency in AI-for-science techniques.

D. **Fund converged knowledge repositories:** Congress should support the creation of structured, AI-ready data platforms that integrate raw results, processed insights, and theoretical knowledge, helping both humans and machines access and reason with scientific data.

E. **Optimize compute resources for research and operations:** Congress should go beyond expanding compute capacity. It should empower the scientific enterprise to balance the computing power used to run AI models with the power needed to advance the AI models themselves. Getting this balance right is essential to sustaining innovation for science. Congress should also promote interoperable frameworks that allow models, tools, and data formats to move seamlessly across research and operational settings, enabling faster adoption of breakthroughs and reducing duplication.

F. **Accelerate where it matters most:** To cut time from discovery to deployment, Congress should strategically blend bold, domain-specific moonshots with cumulative gains across faster-moving fields, targeting acceleration where it delivers the greatest public value.

A. What infrastructure needs to be built to make scientists more productive, and for each type of infrastructure you recommend, what should the funding model be for the construction and operation of that infrastructure?

To significantly accelerate scientific progress in the United States, policymakers should prioritize the development of Self-Driving Labs (SDLs), which are advanced research infrastructure that uses robotics and AI to autonomously design, execute, and interpret experiments. By streamlining the most time-consuming parts of the scientific process, SDLs offer a direct path to higher productivity and faster discovery. They are especially critical in materials science, where progress is currently hamstrung by labor-intensive, trial-and-error experimentation that slows breakthroughs and limits scale.

The challenge in materials science is that researchers still cannot reliably predict how a new material will behave, especially in complex systems or real-world conditions.[2] Unlike molecules in drug development, which can often be simulated with high accuracy, materials interact in messy, nonlinear ways that are difficult to model. Their behavior depends on subtle differences in structure, fabrication, and environmental context, factors that are hard to simulate in advance. As a result, scientists must fabricate and test thousands of physical samples just to discover what works. This demanding process is incredibly inefficient, hard to scale, and a poor use of scientific talent.

SDLs are designed to transform this. These facilities combine robotics, automated synthesis and characterization, and AI-driven decision-making to autonomously design and run experiments. Such labs can dramatically boost productivity by shifting the role of researchers from manual labor to high-level scientific reasoning: designing smarter experiments, interpreting results, and uncovering new principles from large volumes of consistent, high-quality data.

However, despite their immense promise, SDLs are not yet a widespread infrastructure, particularly within materials science. While pioneering prototypes and specialized SDLs exist and are demonstrating capabilities, materials science has not seen the same level of investment in scaling and broadly deploying these systems as related fields like chemistry (e.g., in automated synthesis or high-throughput screening). This is partly because new materials often take a decade or more to reach deployment from discovery to market. While a promising drug may see commercial returns in just a few years, materials breakthroughs face a longer, murkier path to market. That longer timeline discourages significant private investment in advanced infrastructure like SDLs, even though the eventual impact of new materials can be just as transformative.

[2] Sarah Lee, “Computational Materials Science: A New Era.” Number Analytics Blog, June 11, 2025. https://www.numberanalytics.com/blog/new-era-computational-materials-science.

To overcome this investment gap and ensure the construction and operation of SDLs, policymakers should pursue targeted public funding models. Researcher Charles Yang laid out a smart three-pronged strategy last year in the *Federation of American Scientists*.[3] First, an ARPA-E Grand Challenge, modeled on DARPA's initiative for self-driving cars, could jumpstart progress by competitively funding teams to design, build, and demonstrate SDLs capable of delivering genuine breakthroughs in materials discovery. Such a challenge would effectively de-risk early, large-scale investment, while simultaneously showcasing the transformative potential of these labs when fully supported. Second, policymakers could support the creation of Focused Research Organizations (FROs) dedicated to building the modular, open-source hardware and software components that SDLs need to become widely adoptable. Third, the Department of Energy could expand funding for national lab user facilities, where scientists from universities and smaller labs could access SDL infrastructure and gain hands-on experience. This kind of shared infrastructure would accelerate adoption, spread expertise, and help SDLs become part of the scientific mainstream.[4]

B. How do we ensure appropriate design of new scientific workflow models that offload certain tasks to AI while keeping human scientists at the center of the discovery process?

Policymakers are right to care about keeping human scientists integral in the research process, but that doesn't necessarily mean keeping humans at the center of the scientific workflow. In fact, doing so can inadvertently trap scientists in rate-limiting roles that hold back discovery. A more productive approach asks where human insight is most valuable, and where AI, if properly integrated, can take over core scientific tasks to expand the frontier of what's possible. That means rethinking "centrality" not as proximity to the output but as shaping the overall arc of discovery: guiding the inquiry, interpreting results, and asking more ambitious questions.

The early history of numerical weather prediction (NWP) offers a powerful case study in how discovery accelerates when scientists strategically step out of roles that computers can perform more effectively. After World War II, several countries explored computer-generated forecasts, but the United States went further: it redesigned its national forecasting workflow so that computers, not human meteorologists, performed the core task of generating daily forecasts.[5] This wasn't just about automation; it was a fundamental redefinition of roles. The United States was the first country to operationalize computer-generated forecasts, treating them as the authoritative starting point for

[3] Charles Yang, "Automating Scientific Discovery: A Research Agenda for Advancing Self-Driving Labs," Federation of American Scientists, January 31, 2024, https://fas.org/publication/automating-scientific-discovery.
[4] Ibid.
[5] Stanley G. Benjamin et al., "100 Years of Progress in Forecasting and NWP Applications," in *A Century of Progress in Atmospheric and Related Sciences: Celebrating the American Meteorological Society Centennial*, *Meteorological Monographs* 59, no. 1 (2019), https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0020.1.

national weather predictions. This freed human experts to focus on higher-order tasks, such as designing smarter models, interpreting complex results, and explaining forecasts to the public.

The transition wasn't seamless. Early forecasts were often inaccurate due to modeling limitations and messy data.[6] But by operationalizing computer output, U.S. meteorological agencies forced rapid iteration and improvement, and allowed for the pooling of expertise from meteorologists, computer scientists, and mathematicians around the shared challenge of improving the operational system.[7] Over time, this restructuring of the workflow amplified human expertise, pushed the scientific understanding of meteorology to new frontiers, and positioned the United States as a leader in the field for decades. This progress would not have been possible had the process remained strictly human-centered.

This history offers a clear lesson for AI in science today. Preserving the integral role of scientists doesn't mean keeping them at the "center" of the scientific workflow; rather, it means strategically repositioning them where their unique human insights are most valuable. That shift opens the door to faster, more scalable scientific progress.

C. In order to measure the success of ASAP, we need to have objective metrics that measure the speed of scientific innovation. What metrics already exist and what ones need to be created? What information should the federal government have to understand the health and productivity of our innovation ecosystem, and what tools processes, or institutions should be used to do so?

One metric the government should measure is the adoption rates of AI for science among research labs, institutions, and private sector R&D departments, including AI-enabled scientific instruments, software tools like AI/ML platforms, and advanced methodologies. The National Center for Science and Engineering Statistics (NCSES) within the NSF is well-positioned to track this metric in the private sector. It already conducts the Annual Business Survey (ABS) and Business Enterprise Research and Development (BERD) Survey, which include questions about general AI adoption and can be enhanced to ask more granular questions about how the private sector is integrating these technologies into their scientific R&D.[8] NCSES could inquire about the specific types of AI models being adopted for scientific purposes, distinguishing between, for example, general business analytics AI and specialized models for molecular simulation or materials design. The surveys should also probe the depth of integration, determining if AI tools are experimental, used in pilot projects, or fully embedded into core R&D workflows.

---

[6] Ibid.
[7] Charles Yang, “The Early History of Numerical Weather Prediction: The First Compute Arms Race,” (Charles Yang’s website, March 2025), https://charlesyang.io/assets/Supercomputing_and_Weather_Forecasting.pdf.
[8] National Center for Science and Engineering Statistics, Surveys, U.S. National Science Foundation, accessed June 27, 2025, https://ncses.nsf.gov/surveys.

Another metric is workforce proficiency, evaluating the proportion of the scientific workforce skilled in using these cutting-edge AI tools. NCSES is, again, well-placed here. Its National Training, Education, and Workforce Survey (NTEWS) can be enhanced with specific modules.[9] These could ask researchers directly about their proficiency in AI tools, data science, and scientific programming. They could also gauge participation in relevant training programs and their perceived effectiveness. NCSES could also survey employers across sectors to identify emerging skill shortages, ensuring we're training the workforce needed for accelerated science.

D. How can America build the world's most powerful scientific data ecosystem to accelerate American science?

By investing in converged knowledge repositories. These shared digital platforms integrate raw experimental data, scientific theory, and processed insights into coherent, searchable systems. More than mere data warehouses, these repositories are specifically designed to help both human scientists and AI tools to reason with scientific information, not just access it.

The primary challenge is scale. Fields like materials science, biology, and climate modeling generate enormous volumes of data, often petabytes per year, but it's incredibly difficult and time-consuming to make use of this information in its raw form. To unlock its full value, researchers need platforms that can extract relevant patterns, organize raw data into structured formats, such as tables of chemical properties, annotated gene networks, or labeled climate variables, and critically, connect those formats to established scientific principles. These structured forms facilitate easier comparison of findings, more robust hypothesis testing, and the efficient training of AI models.

A highly successful example is the Materials Project, a DOE-backed initiative that provides open access to computed properties for tens of thousands of materials.[10] With over 620,000 registered users and more than 32,000 citations, it demonstrates the profound impact and widespread utility of such a resource.[11] Instead of simply storing raw lab results or simulation outputs, the project uses consistent, physics-based computational methods to calculate the properties of materials in a standardized way, and materials are saved in uniform, searchable formats. This structured approach enables both human scientists and AI models to quickly access and analyze specific material characteristics, without needing to rerun costly, time-consuming experiments or simulations. It not only accelerates discovery but also produces clean, theory-grounded datasets that are ideal for training advanced AI systems.

---

[9] Ibid.
[10] Materials Project, *Next-Gen Materials Project*, https://next-gen.materialsproject.org/
[11] Ibid.

Congress should fund similar efforts across other scientific domains, prioritizing repositories that are FAIR—Findable, Accessible, Interoperable, and Reusable.[12] To be truly effective, these repositories should be built from the outset to serve both human researchers and AI tools. That means incorporating robust APIs, standardized data formats, and integrated tools that allow AI systems to generate hypotheses, test ideas, and refine models from the information. Done right, these platforms won't just store data, they'll actively produce new knowledge, accelerating the next wave of scientific breakthroughs in the United States.

E. What does the U.S. need to do to ensure its researchers have access to enough computing resources to power new breakthroughs?

Accelerating scientific discovery will take more than simply increasing compute capacity. A key challenge is how that compute is allocated, specifically, balancing the resources used to apply AI in scientific workflows with those needed to improve the underlying AI models themselves.

If the United States wants a preview of what happens when that balance goes wrong, it only needs to look at its own decline in weather forecasting. As noted earlier, the nation was once a global leader in applying computing to forecast models, gaining early advantages in both scientific insight and operational performance. But over time, that lead eroded. Today, models developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) routinely outperform its U.S. counterparts across multiple indicators of forecast quality, including overall skill, temperature and wind accuracy, and the reliability of precipitation predictions.[13] The UK Met Office shows similar advantages.

This decline isn't the result of weaker scientific talent in the United States. It's because the United States has underinvested in the computational resources needed to continuously innovate in this domain, while other nations significantly ramped up theirs. Today, the U.S. National Oceanic and Atmospheric Administration (NOAA) dedicates twice as much computing power to daily operations as it does to research, maintaining a research-to-operations ratio of 1:2.[14] In contrast, the ECMWF maintains a ratio close to 5:1 in favor of research. That strategic choice reflects the recognition that long-term forecasting gains come from steady, well-resourced experimentation, and devoting the appropriate computational resources to that end has worked for Europe pulling ahead of the United States.

To be clear, the right balance between research and operational computing won't look the same in every scientific field. As discussed earlier, the pace of discovery in materials science would be well-

---

[12] GO FAIR, *FAIR Principles*, accessed June 27, 2025, https://www.go-fair.org/fair-principles/
[13] Clifford Mass, "NOAA's Forecast Model Has A Drop Out Problem," *Clifford Mass Weather Blog*, May 8, 2025, https://cliffmass.blogspot.com/2025/05/noaas-forecast-model-has-drop-out.html.
[14] Clifford Mass, "The Uncoordinated Giant II," *Bulletin of the American Meteorological Society* 104, no. 4 (2023): E851–E871, doi.org/10.1175/BAMS-D-22-0037.1

served by investing in highly automated self-driving labs, which might require allocating more computational resources toward operational use—running experiments, processing data, and coordinating automation—than toward research compute for things like developing AI that predicts how a new material will perform. The key lesson for ASAP is that it should support the appropriate actors to identify and implement the right computational balance for each field, depending on where the major barriers to progress lie.

Another core issue is computational integration. The United States needs to ensure that scientific advancements can move seamlessly from research labs to operational systems, and across institutions. When one group develops a better method or model, others should be able to adopt it quickly, without having to reinvent the wheel.

Take data assimilation, a computationally intensive technique that improves forecast accuracy by combining model outputs with real-time observations. In Europe, shared software frameworks like the Integrated Forecasting System at ECMWF allow innovations like this to be tested, scaled, and deployed across multiple use cases without friction. But in the United States, forecasting work is spread across multiple federal agencies, including NOAA, the Navy, and the Air Force, each of which operates on siloed hardware and often incompatible software stacks. That fragmentation makes it much harder to share tools or methods. Instead of building on one another's work, agencies often end up duplicating efforts, wasting compute, time, and talent. They don't need to operate the same system, but each system should be designed to work with interoperable AI models, data formats, and tools. That's what enables seamless integration, reduces duplication, and accelerates scientific progress across institutions.

ASAP should therefore support the development and widespread adoption of interoperable technical frameworks tailored to specific scientific fields. While some foundational tools exist, there is a critical need for comprehensive frameworks to provide a common language and a set of core components that researchers and operational centers can build upon, bridging existing silos and ensuring true cross-agency compatibility. This would better allow breakthroughs, whether a new data assimilation technique or an improved physical parameterization, to be directly integrated and scaled across the entire ecosystem without being re-engineered for incompatible systems.

F. In order to cut the time from discovery to deployment by a factor of 10, what changes are needed in the process of scientific innovation, such as in the regulatory ecosystem, scientific funding models, education and workforce pipelines, and the resources that constitute the scientific supply chain?

There are broadly two conceptual pathways to achieving the ambitious goal of reducing discovery-to-deployment time by a factor of 10. One is the "moonshot" pathway, which targets areas with the longest timelines or most entrenched bottlenecks, aiming for bold, breakthrough advances that dramatically shorten the path from scientific insight to real-world use. The other is the "cumulative gains" pathway, which focuses on areas with comparatively shorter timelines or fewer barriers, pursuing a series of smaller improvements across many parts of the scientific innovation ecosystem. While each individual gain may be more modest, their strategic accumulation can collectively yield a tenfold acceleration.

There is a tradeoff between these approaches. The moonshot pathway may deliver deep, transformative impacts in a few high-stakes scientific domains, but those benefits may be concentrated and sector-specific. In contrast, the cumulative gains approach may not achieve such dramatic leaps in any particular field, but can ripple across a broader range of domains, delivering more widespread progress.

Policymakers should remember that cutting discovery-to-deployment time by a factor of 10 is a means to an end, not an end in itself. The goal isn't speed for its own sake, its improving health outcomes, accelerating the rollout of clean energy technologies, and enabling next-generation manufacturing materials. Which pathway to pursue should depend on where acceleration will deliver the greatest public value.