MITRE Response to NITRD "Big Data R&D Initiative" RFI

For additional information about
this response, please contact:

Duane Blackburn
dblackburn@mitre.org
(434) 964-5023

This page intentionally left blank.

Introduction

The MITRE Corporation is a not-for-profit company that runs Federally-Funded Research and Development Centers (FFRDCs) for the U.S. government. FFRDCs are unique organizations sponsored by a government agency to assist with research and development, studies and analysis, and/or systems engineering and integration. Formally established under Federal Acquisition Regulation (FAR) 35.017, FFRDCs meet special, long-term research and development needs that are integral to the mission of the sponsoring agency – work that existing in-house or contractor resources cannot fulfill as effectively. FFRDCs are required to operate in the public interest, be free from conflicts of interest and are operated, managed, and/or administered by universities, not-for-profit organizations, or industrial firms as separate operating units.

MITRE is unique in that our *sole business* is the operation of FFRDCs as well as in the breadth of FFRDC support we provide throughout the federal government. MITRE's seven FFRDCs serve agencies in a variety of areas that impact the public in direct and indirect ways, such as national security; aviation safety and administration; tax administration; homeland security; healthcare; benefits services; cybersecurity partnerships; and other missions.

We are pleased to respond to your request for information regarding big data R&D based on our broad perspective gained from assisting numerous federal entities on big data issues, and from the unique perspective of a systems engineering company that combines a strong technical base with an informed awareness of the larger policy and context in which government operations are conducted.

Vision Statement Thoughts

While MITRE generally agrees with the thoughts conveyed in the draft vision statement, we are concerned that its length and depth will limit its utility. The best vision statements are concise, understandable to outsiders, inspirational to insiders, and clarify the path forward. The draft vision statement is an amalgamation of disparate ideas and platitudes that makes it difficult for readers to truly understand the administration's goals for Big Data R&D. MITRE believes that the administration would be better served with a vision statement similar to the following:

> *Our vision is a public-private innovation ecosystem that fosters collaboration on advancing Big Data technologies that can provide revolutionary capabilities and meaningful information for national priorities in a privacy-sensitive manner.*

With a vision statement such as this, NITRD can ground the vision in practical terms using many of the statements found in the original draft and its succeeding bullets.

Comprehensive Framework Thoughts

Big Data research is taking place throughout the nation – in federal labs, in universities, and (mostly) in the private sector. Interest and investment in Big Data research exists. The predominant remaining needs, therefore, are to organize the research so that it is focused on priority national needs, facilitate collaboration, and encourage knowledge management.

Organize Research. The NSTC, via NITRD, can take a leadership role because it has the necessary breadth of focus combined with the ability to strategically direct federal research funds. We recommend that the first step be an assessment of current capability gaps, followed by a prioritization of those gaps into a published research strategy. Private-sector research aligned with the strategy can be identified and federal investments can help ensure that the highest-priority elements are being adequately addressed.

Facilitate Collaboration. NITRD is uniquely positioned to catalyze broader collaboration among groups of researchers and between researchers and end-users. This is likely best done through a combination of focused workshops and championing collaboration through identifying "use cases" for groups to collectively address. NIST-led efforts such as the NCCOE and NSTIC/IDESG can serve as examples.

Encourage Knowledge Management. Sharing insights and knowledge gained throughout the big data innovation ecosystem will be critical to enhancing the development of this field over the current state. The technical community's typical peer-reviewed paper/poster process moves too slowly to be the predominant method for this field, however, as Big Data researchers require more rapid sharing of discoveries and capabilities gains. Researcher needs would be better served by NSTC/NITRD-enabled events, repositories, and other real-time knowledge management approaches.

Research Needs

MITRE has identified two areas of research need that we feel were not sufficiently addressed in the draft "Vision and Areas of Interest" statement:
- Privacy. While Big Data presents multiple opportunities it also creates privacy concerns, especially when previously separate data sets are combined for uses beyond the original purposes or when de-identified data is combined with information that renders it identifiable. Privacy protections need to be built into system designs from day one rather than considered a "to-be added later" feature. Research into how to best do this needs to be undertaken, and then implemented throughout all subsequent big data research projects.
- Availability of Data Sets. One of the largest costs for Big Data research is developing and maintaining data sets (static, real-time, and synthetic). Efforts to identify and make these available to a broad range of researchers, ideally organized along the identified priorities, would present a significant advancement for the big data innovation ecosystem. Commonly-available datasets, especially when combined with "use case" challenges, have been shown to advance collaboration and spur innovation. Ensuring privacy and IP protection of these data sets will be critical.