

**THE PRESIDENT'S NATIONAL SECURITY
TELECOMMUNICATIONS ADVISORY COMMITTEE**



**BIG DATA ANALYTICS SCOPING REPORT
AUGUST 12, 2015**

**THE PRESIDENT'S NATIONAL SECURITY TELECOMMUNICATIONS ADVISORY
COMMITTEE**
BIG DATA ANALYTICS SCOPING REPORT

I. Overview and Background:

The Government has historically called on both the public and private sectors to harness the power of data in a manner that encourages innovation, promotes national productivity, improves lives, and serves communities, while also emphasizing the importance of individual privacy.

In support of these goals, in 2009, the Administration made tens of thousands of data sets available to the public on data.gov, the central repository for U.S. Government data.

Additionally, in May 2014, the White House released the *Big Data: Seizing Opportunities, Preserving Values* report, which examined how big data could “alter the relationships between Government, citizens, businesses, and consumers.” Overall, the report “focuses on how the public and private sectors can maximize the benefits of big data while minimizing its risks,” as well as “identifies opportunities for big data to grow [the U.S.] economy, improve health and education, and make [the] Nation safer and more...efficient”.¹

In addition to its own big data efforts, the Government is exploring how private sector data sets and infrastructural resources can be utilized in support of its national security and emergency preparedness (NS/EP) activities. As a result, the Government now has the opportunity to learn from industry best practices as it seeks to create policies that will allow it to: (1) identify current and emerging big data sets within the public and private sectors that could be used in support of NS/EP functions; (2) select and/or develop models that will further encourage information sharing within Government and between Government and the private sector; and (3) access big data sets to support NS/EP capabilities, when appropriate. While pursuing this opportunity, Government must also be mindful of the policy challenges, particularly those involving privacy, associated with leveraging big data within the NS/EP context.

Therefore, in February 2015, the Executive Office of the President requested that the President’s National Security Telecommunications Advisory Committee (NSTAC) examine how big data analytics (BDA) could enhance national NS/EP capabilities.² In response, the NSTAC Big Data Analytics Subcommittee was formed in March 2015.

II. Estimated Time Frame and Priority:

The NSTAC plans to complete its final report by May 2016.

III. Value in Researching Issue:

While the data generated by individual and organizational users is rapidly increasing, the associated costs to collect, store, and process data are declining. The use of new data sources, such as sensors and various networked devices, is accelerating the growth of large information stores. These sensors and devices have demonstrated the ability to generate ever-greater volume, velocity, and variety of data. To effectively process and leverage big data, technology has needed to constantly evolve to meet user requirements for these data sets. As a result,

¹ Executive Office of the President. *Big Data: Seizing Opportunities, Preserving Values*. May 1, 2014. Available at: https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf.

² In the absence of authoritative definitions of big data, BDA, and related phrases, the NSTAC will continue to use the term BDA throughout the scoping and research phases.

current big data infrastructure and analytical tools offer powerful opportunities to unlock previously inaccessible insights from new and existing big data sets. In light of these advancements, new opportunities may exist to utilize BDA in support of Government's NS/EP functions (e.g., disaster planning and response, cybersecurity, analytics, and communications/network resiliency).

IV. Approach:

The NSTAC will examine the implications of BDA on Government's NS/EP functions. This examination will focus on three fundamental areas: (1) identifying and optimizing the use of current data streams, including those operationalized or designed for other specific uses; (2) creating processes to discover evolving data sets that can be used in support of NS/EP functions; and (3) identifying challenges, risk areas, and appropriate mitigating activities for those risks. As BDA continues to proliferate, mitigation efforts will likely require significant attention to ensure that the implicit challenges associated with this information (e.g., security, privacy, accuracy, and validity) do not negate its benefits for the Government's NS/EP functions.

The NSTAC will scope its investigation by studying NS/EP use cases and will map each use case to the lifecycle commonalities seen in the implementation of big data. The NSTAC will then take the lessons learned from each case study and structure its recommendations around the three fundamental areas listed above. These use cases could focus on, but are not limited to, the study of BDA and its use surrounding:

- A natural disaster;
- A kinetic terrorist attack; and
- The cyber resiliency of critical infrastructure.

The use of the data lifecycle will allow for a linear and structured approach that can be applied to the big data use cases listed above. The following steps form the primary elements of the data lifecycle and are accompanied by questions to illustrate the type of issues that the NSTAC will examine at each stage.

1. Data creation and collection:

- a. What type of data is being collected, by whom, and for what purpose?
- b. What rules and regulations constrain original data collection?

2. Data identification, availability, and quality for BDA:

- a. How can analysts identify or how do data owners make available public and private data sets for BDA?
- b. How is the data transmitted/transferred from its original source?
- c. Should there be different use practices for data collected from private versus public sources?
- d. What are the privacy and security issues related to use of data for BDA?
- e. How does data need to be cleansed or processed to make it usable for BDA?

3. Data analytics:

- a. How do constraints on the data affect what and how BDA can be conducted?
- b. Who has the rights to conduct BDA on data that was originally collected for other purposes?

- c. What legal issues arise when BDA is conducted on data originally collected for another purpose?
- d. What are the objectives of the analytics (e.g., reactive, preventive, forensic, or predictive)? Do particular objectives create distinct policy issues?
- e. What is the operational model for conducting BDA? How do Federal entities build, source, or share analytic platforms for conducting BDA?
- f. What are the privacy and security issues that arise from:
 - i. The transfer of large data sets from the original data owner to the entity conducting the analytics?
 - ii. The integration of data from multiple sources?
 - iii. How and by whom the data is analyzed (e.g., on premise, Federal shared-service, or commercial cloud)?

4. Use of analytic outputs:

- a. How are BDA activities integrated with existing organizational processes and mission activities to ensure timeliness, impact, and ease of use?
- b. Who owns the rights to BDA derived data and the analysis results?
- c. Have precautions been taken to ensure the analytic products are nondiscriminatory and do not infringe upon individual privacy?
- d. How are analytic results shared and/or published to provide actionable support for decision making?
- e. How are analytic results shared, published, and operationalized both within an organization and with external stakeholders?
- f. What domain models or standards can effectively facilitate the sharing of analytic outputs?

5. Data management, storage, retention, and disposal:

- a. How should retention and disposal of the data be handled?
- b. How should retention and disposal of the analysis results be handled?
- c. Who has access to the data and how is it shared?
- d. What type of protection policies should be in place for stored data?
- e. What issues need to be addressed to facilitate data sharing within Government and between Government and the private sector?
- f. What risks are involved with the storage and retention of big data sets?
- g. What are the unique considerations of retaining data for extended periods of time?

For each general use case category, the NSTAC can focus its study on multiple specific examples. From these examples, the NSTAC will then select a specific demonstrative use case or scenario from each category when forming its policy recommendations for the President. For instance, in the natural disaster category, the NSTAC may develop a scenario discussing the use of BDA before, during, and after a significant earthquake that has occurred in the Western United States. This structure will allow the NSTAC to shape its policy recommendations from individual steps in the data lifecycle in relation to its three fundamental research areas.

The NSTAC does not intend to produce a technical report on BDA; however, the inherent nature of the topic demands that the NSTAC's study include an examination into how policy can address the technological, standardization, and infrastructure requirements associated with effectively using big data to support Government's NS/EP capabilities. Moreover, while the study will also identify the privacy and civil liberty implications of using BDA, the NSTAC's

main focus will be to investigate and recommend policies relative to the implications of BDA on the Government's NS/EP functions.

To perform the research and develop recommendations, the NSTAC will:

- Receive briefings from subject matter experts (SMEs) who are engaged with BDA. The identified briefers will provide insight into applicable technical issues, lessons learned, best practices, and/or research activities. At a minimum, briefers should include representatives from leading organizations and related projects from Government, industry, the nonprofit sector, and academia.
- Review relevant academic literature and current BDA research studies.
- Review industry's best practices and capabilities for possible relevancy/portability.
- Evaluate current efforts and policy documents, as well as previous NSTAC reports, for applicability.

Nothing included within this scoping document is intended to constrain the NSTAC from investigating all necessary aspects of the issue and correcting any implicit or explicit oversights that may appear in this brief, anticipatory scoping report.

V. Proposal:

Recommend that the NSTAC, through the BDA Subcommittee, continue its study into the implications of BDA for Government's NS/EP functions and produce a report for the President's review and consideration.

VI. Schedule:

- **August 12, 2015:** NSTAC members deliberate and vote on the NSTAC BDA Scoping Report in order to continue the NSTAC BDA Subcommittee's investigation into the implications of BDA on Government's NS/EP functions.
- **August 18, 2015:** Hold the first research subcommittee meeting.
- **August – November 2015:** Identify and invite relevant SMEs to brief the subcommittee.
- **October 2015:** Invite the NSTAC members to participate in a subcommittee meeting to discuss emerging themes and topics that have arisen during the research phase.
- **November 10, 2015:** Provide an update to the NSTAC members at the November 2015 NSTAC Meeting.
- **November 10, 2015 – January 2016:** Continue research, receive briefings from SMEs, and begin drafting the final report.
- **February 2016:** Provide an update to the NSTAC members during the Member Conference Call in a manner that will allow time for member comment, discussion, and feedback.
- **February or March 2016:** Invite the NSTAC members to participate in a subcommittee meeting to discuss potential recommendations to the President regarding the implications of Government using BDA in support of NS/EP functions.
- **April 2016:** Finalize the draft report.
- **May 2016:** Present the final report at the 2016 NSTAC Meeting for deliberation and vote.