RFI Response for National Science Foundation (NSF) – National Big Data RFI Initiative

1. Who you are—name, credentials, organization, contact information; your experience working with big data and your role in the big data innovation ecosystem

CSC is a global leader of next-generation IT services and solutions. Our mission is to enable superior returns on clients' technology investments through best-in-class industry solutions, domain expertise and global scale. CSC brings our broad and deep expertise, in concert with a proven track record of extracting, processing and implementing IT and data solutions across a variety of clients. CSC has successfully delivered a multitude of complex data integration projects for federal and state government agencies, hospitals, and global healthcare institutions (e.g. United Kingdom National Health Service program) and Defense and Intel customers.

Point of Contact (POC):
Yvonne Chaplin
Big Data & Analytics Brand Manager, Computer Sciences Corporation
3170 Fairview Park Drive
Falls Church, VA 22042
C: 571-294-9875
email: ychaplin@csc.com

2. What are the gaps that are not addressed in the Visions and Priority Actions document?

1) **Organizational Transformation:** To fully realize the value of Big Data, organizations will need to take action based on the insights that the data provide. For example, data analysis with the purpose of enhancing the customer experience may lead to recommended changes in business process, data sources, infrastructure, architecture, and organizational structures.

2) **Use Cases:** We recommend the National Big Data Initiative identify specific use cases to which big data solutions may provide remedies. The focus on specific problems to solve would ensure practical use – rather than building a solution in search of a problem.

3) **Data Governance**: Data governance within and across organizations is key to ensuring long-term sustainability and access. Data governance agreements that articulate who can access the data at what level, privacy considerations, and de-identifying personal information will help to break down the "stovepipes" between organizations in a secure manner.

3. From an interagency perspective, what do you think are the most high impact ideas at the frontiers of Big Data research and development?

1) **Public safety:** Enhanced forecasting of the extent and impact of natural and man-made disasters and improved status will continue to help optimize the deployment of resources to support and protect the population.

2) **Electronic health records (EHRs)**: EHRs coupled with new analytics tools present an opportunity to mine information for the most effective outcomes across large populations.

Using carefully de-identified information, researchers can look for statistically valid trends and provide assessments based upon true quality of care.

3) **Health sensors:** Used in the hospital or home, these sensors can provide continuous monitoring of key biochemical markers, performing real time analysis on the data as it streams from individual high-risk patients to a HIPAA-compliant analysis system. By providing alerts to a health anomaly or pending critical event, sensors have the potential to extend and improve the quality of millions of citizens' lives.

4) **Education:** In-depth tracking and analysis of on-line student learning activities – with fine grained analysis down to the level of mouse clicks – can help researchers to ascertain how students learn and the approaches that can be used to improve learning. This analysis can be done across thousands of students and can inform the development of courses and teaching approaches to reflect the information gleaned from the large scale analysis.

4. **What new research, education, and/or infrastructure investments do you think will be game-changing for the big data innovation ecosystem?**

1) **Research:** We recommend continued investment and support of Open Source projects to promote speed and collaboration among a community of Big Data innovators. In addition, we are seeing an explosion of social media data that is publicly available. Enterprises and Government have incredible opportunities to fuse their own data with this public social media data to draw intelligence.

2) **Education and training:** Data scientists will require a multi-disciplinary skill-set combining business, mathematics, statistics and engineering. Data scientists are not typically technologists, and often have great difficulty manipulating the complex technologies that span the big data spectrum. R&D should focus on bridging this gap by developing tools (one example is open source Amino) to simplify this complexity for the analyst.

3) **Infrastructure:** We recommend that infrastructure R&D address delivery of information through varied, multiple, and concurrent channels and mobile devices.

5. **How can the federal government most effectively enable new partnerships, particularly those that cross sectors or domains?**

Data governance agreements and security standards will be critical in enabling disparate parties to form partnerships, providing a degree of assurance that privacy and sensitive data will be appropriately accessed and protected.

6. **A short explanation of why you feel your contribution/ideas should be included in the strategic plan (examples where appropriate)**

CSC's response and recommendations are based on practical "real-world" experience of applying Big Data strategies and technologies and developing solutions for multiple clients in the public and private sectors. Examples are briefly cited in our response to the first question above.