**NIST Big Data Public Working Group (NBD-PWG)**

**NBD-PWD-2015/M0374**

**Source: NBD-PWG**

**Status: Draft**

**Title: Prioritize NBD Version 2 Works – Based from Version 1 Future Works**

**Author: NBD-PWG V1.0 Documents**

## Vol.1 Definitions Future Work

Subsection focus: Discuss the future updates that are planned for this Volume. We should mention that we want to harmonize with JTC1 efforts.

This volume is a first attempt of the working group to provide order and clarity to an emerging and rapidly changing field. Given the breadth of the types of data, the fields of study, the technologies, and the techniques, it has been a challenge to distill from the many viewpoints a consistent core set of definitions to begin to frame the discussion and understanding of this new paradigm. As the field matures, this document will also need to mature in order to accommodate new innovations in the field. To ensure the concepts are accurate, future working group tasks will consist of defining the different patterns of communications between Big Data resources to better clarify the different approaches being taken. This will provide further clarity to the broader community seeking to understand the field.

A number of other efforts are underway to address the community’s needs in the Big Data space. ISO/JTC 1 formed a Study Group on Big Data to identify areas where current standards efforts could be applied to big data, and areas that might need additional attention. Four meetings were held around the world, and input was solicited from all ISO/JTC participating national bodies. The NIST PBDWG both contributed to this discussion through the US National body INCITS Big Data Ad Hoc, and benefited from that process. The ISO/JTC response to the Study Group report was to provide some guidance to existing committees, and to form a JTC 1 Working Group 9 on Big Data. Going forward the NIST working group will want to refine this work to remain in sync with the JTC and other working group efforts.

The NPBDWG is currently collaborating with the Research Data Alliance, which is exploring the interfaces between components. We have an initial representation in Section 4 Big Data Engineering patterns, but that section will improve through the additional work in this partnership.

The NPBDWG has been in communication with the TCP-xHD Big Data Committee, which been addressing system benchmarks. Information from their efforts will be included in future versions of this report.

<Mention CSC Taxonomy and S&P?>

## Vol. 2 Taxonomies Future Work

As mentioned in the previous section, the working group is continuing to explore the changes in both Management and in Security and Privacy. As the changes in the activities within these roles are better clarified, the taxonomy will be developed in these areas. In addition, a fuller understanding of Big Data and its technologies should focus on the interactions between the characteristics of the data and the desired analytics in both technique and time window for performance. These characteristics drive the application, and the choice of tools to meet the system requirements. An investigation of the interfaces between the data characteristics and the technologies is a continuing task within the definitions and taxonomy subgroup, and the reference architecture subgroup. Finally, societal impact issues have not been fully explored. There are a number of over-arching issues in the implications of Big Data, such as data ownership and data governance that need more examination.

## Vol. 3 Use Cases & Requirements Future Work

Additional work on the use cases is in progress, some of which is included in papers submitted to the ACM workshop attached to this process. Future work includes an identification of general features or patterns and a classification of use cases by these features. The classification leads to suggested classes of software models and system architectures. We intend a more detailed analysis of reference architecture based on sample codes that are being implemented in a class.

Other future work could include collection and classification of additional use cases where areas, including Government Operations, Commercial, Internet of Things, and Energy, would benefit from additional entries. Additional information on current or new use cases may become available, including associated figures. In future use cases, more quantitative specifications could be made, including more precise and uniform recording of data volume. Further requirements analysis can be performed now that the reference architecture is more mature.

## Vol. 4 Security & Privacy Future Work

The NBD-PWG Security and Privacy Subgroup plans to further develop several topics for the subsequent version (Version 2) of this document. These topics include the following:

* A closer examination of other templates. These templates may be adapted to the Big Data security and privacy fabric to address gaps in Version 1 and to bridge the efforts of this Subgroup with the work of others.
* Further developing the Security and Privacy Taxonomy
* Enhancing the connection between the Security and Privacy Taxonomy and the NBDRA components.
* Developing the connection between the Security and Privacy fabric and the NBDRA.
* Expanding the privacy discussion within the scope of Volume 4
* Exploring governance with respect to security and privacy
* Mapping the identified security and privacy use cases to the NBDRA
* Contextualize the content of Appendix B in the NBDRA

## Vol. 5 Architecture White Paper Survey – Done

## Vol. 6 Reference Architecture Future Work

There are three stages for developing the NIST Big Data Reference Architecture (NBDRA). The developments in each of these three stages will be reported in three versions of Volume 6, as follows: (1) Identify the common, key reference architecture components, (2) Define general interfaces between the NBDRA key components, and (3) Validate the NBDRA by building Big Data general applications through the general interfaces.  This document (Version 1) will present the overall RA key components with high-level description and functionalities.

Version 2 activities will focus on how to define general interfaces between the RA key components by:

* Pick few use cases from the 62 (51 general +11 SnP) submitted or other meaningful use cases
* Work with domain expert to identify workflow and interactions from the System Orchestrator to rest of the RA components
* Implement small scale, manageable and well-defined confined environment and model interaction between key components.
* Aggregate the common data workflow and interaction between key components and package them into general interfaces

Version 3 activities will focus on RA validation: Can the defined RA interface build general Big Data applications? The strategy will be:

* Implement the same set of use cases used in Version 2 by using the defined general interfaces
* Identify and implement few new use cases outside the Version 2 scenarios
* Enhance general interfaces through lesson learned from the above implementations

Note: General interface developed from Version 3 will not be perfect but rather, it is offered as a starting point to get further refinement from any interested parties.

## Vol. Standards Roadmap Future Work – TBD