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

**NBD-PWD-2015/M0453**

**Source: NBD-PWG**

**Status: Draft**

**Title: NBD-PWG V2 Activities, Deliverables, and Timelines, rev-1**

**Author: NBD-PWG Subgroup Co-Chairs**

**NBD-PWG V2 Activities, Deliverables, and Timelines**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Activities**  **NBD**  **Documents** | **V1**  **RA**  **Components**  **[09/16/15]** | **V2**  **RA**  **Interfaces**  **[06/15/16]** | **White Paper**  **RA**  **Best Practices**  **[ongoing]** | **V3**  **RA**  **Validation**  **[03/15/17]** | **ISO/IEC**  **JTC 1/WG9**  **[12/31/17]** |
| **Overview &**  **Vision** | -- | -- | **Based from the NIST six use cases (M0399)**  Activities and Functional exercise (M0437)   1. UC#1: Fingerprint Matching 2. UC#2: Human Face Detection from Video   Ann, [date]   1. UC#3: Live Twitter Analysis   PW, Russell, Gretchen, [date]   1. UC#4: Big data Analytics for Healthcare   Russell, Mark [date]   1. UC#5: Spatial Big data/Spatial Statistics/GIS 2. UC#6: Data Warehouse and Data Mining   Russell, Korin, Ann, Sharaf [date]   1. Big Data for Security   Sharaf [date]   1. Security and Privacy   Cavan, [date]   1. Others…   Implementation exercise   1. UC#1: Fingerprint Matching   <names>, [date]   1. UC#2: Human Face Detection from Video   <names>, [date]   1. UC#3: Live Twitter Analysis   <names>, [date]   1. UC#4: Big data Analytics for Healthcare   Wo, Nancy, TC69, [12/30/15]  Ian and his students, [date]   1. UC#5: Spatial Big data/Spatial Statistics/GIS   <names>,[date]   1. UC#6: Data Warehouse and Data Mining   <names>, [date]   1. Others… | -- | ISO 20547-1 [TR]  Framework & Process  1st WD: 12/04/15  1st ED: 10/01/15 |
| **Vol. 1**  **Definitions** | Done | 1. Governance, ownership   Ann, Jim, [date]   1. Provenance, curation   Jim [date]   1. Others… | Wait for V2 | ISO 20546 [IS]  Overview &  Vocabulary  2nd WD – 12/04/15  1st WD – 07/20/15  1st ED – 04/17/15 |
| **Vol. 2**  **Taxonomies** | Done | 1. Harmonize with RA components   Nancy, [date]   1. Others… | Wait for V2 | -- |
| **Vol. 3**  **Use Cases &**  **Requirements** | Done | 1. Seek use cases using new UC template   Geoffrey, Piyush, [date]   1. Others… | Wait for V2 | ISO 20547-2 [TR]  UC & Derived Reqs.  1st WD: 12/04/15  1st ED: 10/01/15 |
| **Vol. 4**  **Security &**  **Privacy**  **Fabric** | Done | 1. Universal SnP Taxonomy   Arnab, Mark, [date]   1. Explore NIST IR 8062,   Mark, [date]   1. Others… | Wait for V2 | ISO 20547-4 [IS]  -- Same Title --  1st WD: 12/04/15  1st ED: 10/01/15 |
| **Vol. 5**  **White Paper**  **Survey** | Done |  | -- | -- |
| Wait for V2 | ISO 20547-3 [IS]  -- Same Title --  1st WD: 12/04/15  1st ED: 10/01/15 |
| **Vol. 6**  **Reference**  **Architecture** |
| Done | 1. Enhance RA components description   Dave, [date]   1. Establish preliminary RA interfaces 2. Dave, [date]Others… |
| **Vol. 7**  **Standards**  **Roadmap** | Done | 1. Extend related SDOs listing   Russell, [date]   1. Enhance gap analysis how to enable RA   Russell, [date]   1. Others… | Wait for V2 | ISO 20547-5 [TS]  -- Same Title --  1st WD: 12/04/15  1st ED: 10/01/15 |

**Note: TR – Technical Report, IS – International Standard, ED – Editors Draft, WD – Working Draft**

**NBD-PWG V2 Works**

1. From Vol. 1 Definitions, Section 1.5 Future Work
2. Defining the different patterns of communications between Big Data resources to better clarify the different approaches being taken;
3. Updating Volume 1 taking into account the efforts of other working groups such as International Organization for Standardization (ISO) Joint Technical Committee 1 (JTC 1) and the Transaction Processing Performance Council;
4. Improving the discussions of governance and data ownership;
5. Developing the Management section;
6. Developing the Security and Privacy section; and
7. Adding a discussion of the value of data.
8. From Vol. 2 Taxonomies, Section 1.5 Future Work

The Subgroup is continuing to explore the changes in both Management and in Security and Privacy. As changes in the activities within these roles are clarified, the taxonomy will be developed further. In addition, a fuller understanding of Big Data and its technologies should consider the interactions between the characteristics of the data and the desired methods in both technique and time window for performance. These characteristics drive the application and the choice of tools to meet system requirements. Investigation of the interfaces between data characteristics and technologies is a continuing task for the NBD-PWG Definitions and Taxonomy Subgroup and the NBD-PWG Reference Architecture Subgroup. Finally, societal impact issues have not yet been fully explored. There are a number of overarching issues in the implications of Big Data, such as data ownership and data governance, which need more examination. Big Data is a rapidly evolving field, and the initial discussion presented in this volume must be considered a work in progress.

1. From Vol. 3 Use Cases & Requirements, Section 1.5, Future Work
2. Identify general features or patterns and a classification of use cases by these features.
3. Draw on the use case classification to suggest classes of software models and system architectures.[[1]](#endnote-1),[[2]](#endnote-2),[[3]](#endnote-3),[[4]](#endnote-4),[[5]](#endnote-5)
4. A more detailed analysis of reference architecture based on sample codes that are being implemented in a university class. [[6]](#endnote-6)
5. Collect benchmarks that capture the “essence” of individual use cases.
6. Additional work may arise from these or other NBD-PWG activities. Other future work may include collection and classification of additional use cases in areas that would benefit from additional entries, such as Government Operations, Commercial, Internet of Things, and Energy. 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. In addition, further requirements analysis can be performed now that the reference architecture is more mature.
7. From Vol. 4 Security & Privacy, Section 1.5, Future Work
8. Examining closely other existing templates[[7]](#footnote-1) in literature: The templates may be adapted to the Big Data security and privacy fabric to address gaps and to bridge the efforts of this Subgroup with the work of others;
9. Further developing the security and privacy taxonomy;
10. Enhancing the connection between the security and privacy taxonomy and the NBDRA components;
11. Developing the connection between the security and privacy fabric and the NBDRA;
12. Expanding the privacy discussion within the scope of this volume;
13. Exploring governance, risk management, data ownership, and valuation with respect to Big Data ecosystem, with a focus on security and privacy;
14. Mapping the identified security and privacy use cases to the NBDRA;
15. Contextualizing the content of Appendix B in the NBDRA; and
16. Exploring privacy in actionable terms based on frameworks such as those described in NISTIR 8062[[8]](#endnote-7) with respect to the NBDRA.
17. From Vol. 5 White Paper Survey, No Future Work for now
18. From Vol. 6 Reference Architecture, Section 1.5, Future Work
19. Select use cases from the 62 (51 general and 11 security and privacy) submitted use cases or other, to be identified, meaningful use cases;
20. Work with domain experts to identify workflow and interactions among the NBDRA components and fabrics;
21. Explore and model these interactions within a small-scale, manageable, and well-defined confined environment; and
22. Aggregate the common data workflow and interactions between NBDRA components and fabrics and package them into general interfaces.
23. From Vol. 7 Standards Roadmap, Section 1.5, Future Work
24. Continue to build and refine the gap analysis and document the findings;
25. Identify where standards may accelerate the adoption and interoperability of Big Data technologies;
26. Document recommendations for future standards activities; and
27. Further map standards to NBDRA components and the interfaces between them.

1. Shantenu Jha, Judy Qiu, Andre Luckow, Pradeep Mantha, and Geoffrey C. Fox, “A Tale of Two Data-Intensive Approaches: Applications, Architectures and Infrastructure, in 3rd International IEEE Congress on Big Data Application and Experience Track,” *Cornell University Library,* June 27- July 2, 2014*,* <http://arxiv.org/abs/1403.1528>. [↑](#endnote-ref-1)
2. Judy Qiu, Shantenu Jha, Andre Luckow, and Geoffrey C. Fox, “Towards HPC-ABDS: An Initial High-Performance Big Data Stack,” *Indiana University,* August 8, 2014. <http://grids.ucs.indiana.edu/ptliupages/publications/nist-hpc-abds.pdf>. [↑](#endnote-ref-2)
3. Geoffrey Fox, Judy Qiu, and Shantenu Jha, “High Performance High Functionality Big Data Software Stack, in Big Data and Extreme-scale Computing (BDEC),” *Indiana and Rutgers Universities*, 2014. <http://www.exascale.org/bdec/sites/www.exascale.org.bdec/files/whitepapers/fox.pdf>. [↑](#endnote-ref-3)
4. Geoffrey C. Fox, Shantenu Jha, Judy Qiu, and Andre Luckow, “Towards an Understanding of Facets and Exemplars of Big Data Applications,” *Indiana University,* July 20, 2014. <http://grids.ucs.indiana.edu/ptliupages/publications/OgrePaperv9.pdf>. [↑](#endnote-ref-4)
5. Geoffrey Fox and Wo Chang, “Big Data Use Cases and Requirements,” *Indiana University,* August 10, 2014. <http://grids.ucs.indiana.edu/ptliupages/publications/NISTUseCase.pdf>. [↑](#endnote-ref-5)
6. Geoffrey Fox. “INFO 590 Indiana University Online Class: Big Data Open Source Software and Projects,” *Indiana University,* 2014 [accessed December 11, 2014], <http://bigdataopensourceprojects.soic.indiana.edu/>. [↑](#endnote-ref-6)
7. There are multiple templates developed by others to adapt as part of a Big Data security metadata model. For instance, the subgroup has considered schemes offered in the NIST Preliminary Critical Infrastructure Cybersecurity Framework (CIICF) of October 2013, <http://1.usa.gov/1wQuti1> (accessed January 9, 2015.)). [↑](#footnote-ref-1)
8. DRAFT Privacy Risk Management for Federal Information Systems

   http://csrc.nist.gov/publications/drafts/nistir-8062/nistir\_8062\_draft.pdf [↑](#endnote-ref-7)