**Minutes of NBD-PWG/Subgroups Joint Meeting on September 4, 2013**

Wo Chang opened the meeting. New members are continuing to join the NIST Big Data Working Group. About 50 new members have joined in the past few weeks, since the last meeting. The charter and overview of activities of the sub-groups was provided. See 2013\_09\_04NBD-PWG-Subgroups\_progress\_pptx (<http://bigdatawg.nist.gov/_uploadfiles/M0212_v1_3404467806.pdf>).

The rest of the meeting will review the work done so far and also provide updates. Each subgroup will share their activities

Taxonomy subgroup. Presentation by Nancy Grady

* Are discussing access and transformations of external vs internal data
* How to represent overarching issues related to data, which haven’t changed much from what has always been done (before big data), e.g. governance, etc.
* The taxonomy group is working with the Architecture group to reconcile terminology.
* The taxonomy document is available in MindMeister.com which provides sharing capability with dynamic tracking. For those interested in updating, please send email to Wo Chang.

Requirements subgroup. Presentation by Geoffrey Fox

* They have 42 different use cases so far. Another three new use cases have been contributed recently. Many of the use cases with complete information are on the science applications side.

Privacy and Security subgroup. Presentation by Arnab Roy

* The subgroup has been making good progress with their document.
* An example was provided of how security requirements are derived from applications requirements using the Cargo Shipping Example in the Privacy subgroup document.
* The subgroup will similarly go thru all use cases from the Requirements subgroup to look for commonality of security/privacy requirements and also any new requirements generated by a given use case.

Architecture subgroup. Presentation by Orit Levin

* The RA on-going document (M0100) on proposed big data architecture(s) was discussed, with the current architecture diagram (2 alternatives).
* There is opportunity to provide input about alternative big data architectures. See M0151 for the template for providing this input. The template has three items that need to be addressed: (a) general description, (b) architecture model, and (c) key components and their functionalities.
* Regarding the use cases from the Requirements subgroup, Wo mentioned that the architecture needs to address the generalized requirements across all use cases, rather than focus on specific, individual use cases. Convergence is needed and being sought in the Architecture subgroup, between the two architecture diagrams, by answering the open questions.

Carl Buffington: Technology RoadMap

* Want this document to be a rollup of work from the various other subgroups, so there is a dependence here.
* The group is on track to produce a document for Sept 30th meeting.

Open Discussion

* Wo: went thru chat comments. Most items were remarks/comments.
* Nancy: Discussion of taxonomy document. Have not put up the updated document from Monday (9/2). Will do so—everyone is requested to provide their comments.
* Geoffrey: Use case documents has been updated now to include 45 use cases. If there are any more use cases, please try to get them to Geoffrey by today (9/4).
* Arnab: need to synch security document with architecture document.
  + Taxonomy and Reference Architecture groups are meeting on 9/5 to reconcile materials. Security will join the discussion.
* Orit: reviewed the M0151 document, to see what the templates look like for the Big Data Architecture alternative proposals.
  + Each architecture will have:
    - Description;
    - Architecture Model, e.g. a workflow/dataflow model, from step to step; or a Specification-based approach; etc. Can learn what is a good model for specifying the architecture itself.
    - Key Feature Components
  + M0123 is the regular architecture document
* Carl: looking for a brief overview, summary from other groups in terms of the essence / takeway for the Tech Roadmap.
* Group Discussion re big data as a “data type”.
  + How big does data have to be to be “big”? Is it only volume, e.g. GB, TB, or? It depends, because what is considered “big” changes. Don’t want a definition that may become obsolete soon. Is there a way to quantify the notion of “big data”? A function of dataset size, # of cores, memory size, etc. (Ed-what about algorithmic complexity and scalability?)

Summary of action items for document exchanges:

1. Def. & Tax. (DnT): supply high-level Big Data Definitions and Taxonomies (mainly Section 2) to RA, SnP, and TR.
2. Requirements (Reqs): supply high-level generalized extracted requirements (Section 3 – only the aggregated requirements; so far we have done the first eight use cases but will complete the rest soon) to RA and TR.
3. Security & Privacy (SnP): supply high-level Taxonomies and Arch. to DnT, RA, and TR.
4. Ref. Arch (RA): supply high-level RA key components and interfaces to DnT, SnP, and TR.
5. Tech. Roadmap (TR): incorporate DnT, Reqs, SnP, RA into Section 2 to 5 of the Tech. Roadmap document.