**NIST BD joint subgroup meeting minutes for Dec 2, 2013**

Prepared by Sanjay Mishra, Verizon

**Scope**

* Recap Current Deliverables and Share Future Activities

Current Deliverables:

* + V1.0 Documents for Requirements and Sec. & Privacy – for technical editing
  + V1.0 Documents for Def. & Tax., Ref. Arch., and Tech. Roadmap – need harmonization
  + Schedule: every Monday, 1:00PM – 4:00PM EDT till Dec. 23, 2013
  + Goal: finalize all V1.0 draft documents (ready for technical editing) **by Dec. 23, 2013**
  + Reflector: please use [bigdatadeftax@nist.gov](mailto:bigdatadeftax@nist.gov), [bigdataarch@nist.gov](mailto:bigdataarch@nist.gov), and/or [bigdataroadmap@nist.gov](mailto:bigdataroadmap@nist.gov) for discussion

**Monday, December 02, 2013**

1. **Action Items**
2. Nancy and Orit to continue to collaborate and update section 3 of RA incorporating from Definition and Taxonomy. Specific updates to include review of Levels 1 and 2 (subsequently Levels 3 and 4) and Framework Provider and System Management
3. System Management needs (section 4.1) updated (Owner: Felix Njeh)
4. Need a synch up with Security (Owner: Orit)
5. All participants should review mind-map to see what is missing, from actors, and activities perspective. (Owner(s): All) For accessing mind-map, please send email to Wo.
6. Security and Privacy – Smaller sub-teams are working on document updates and a version is expected to be available in two weeks
7. US INCITS Big Data Study Group and ISO/IEC JTC 1 Big Data Study Group general information and meeting logistics
   1. Wo informed this study group will meet 3 times over the 12 month period. Once in US, followed by Europe and China. Wo is working on securing hosting sites and will likely include University settings.
   2. The first meeting is tentatively planned for 4-days starting Jan 28 and through Jan 31.
   3. There is no fee for working on the Study Group
   4. Additional details to be published by Wo
   5. Anyone interested in collaborating on this, please reach out to Wo (wchang@nist.gov)

3. **Next meeting: Dec. 9, 1:00PM - 4:00PM EST**

**Definition and Taxonomy**

1. **Definition & Taxonomy**
   * Nancy reviewed Mind-map (Roles, activities, components and sub-components) and D&T alignment with the section 3 of the Reference Architecture.
   * (PW) Do you think we can define 'Data Ownership'....or is that too much....?
   * (John Dodd) is registration or data registry the same as data availability publication, or is there a "consent data form" on primary and secondary uses of the data (for access rights and responsibilities)
     + Nancy Added, Access Rights Management to Mind-map
     + Nancy also added a place holder for compression to ease data distribution (per William Vorhiess)
2. Orchestrator

* Should “Audit” be added here? (Per Nancy, can be covered under Data Provider contracts and SLAs)

1. Application Provider

* Added Human-in-the-loop rapid Analytics (discovery, hypothesis, hypothesis testing, alternate methods) within Data Analytics Processes

1. Framework Provider
   * + Need some more work

**Security & Privacy (Arnab Roy)**

1. Arnab reported work in progress and will be presenting soon
   1. Smaller teams are working currently and are at different level of maturity. Arnab expects a version to be available in about 2 weeks.
2. Objective is to link S&P with Reference Architecture

**Reference Architecture**

Changes to the document

1. Orit reported specific updates to section 2, 3 and 4 have been made (need to enable track changes to easily identify changes)
2. Section 2 last paragraph: if this paragraph expands with more substantial text, it would be good to have its own section and called it Abstraction Services. The idea is to demonstrate to reader how the RA key components interact with each other.
3. Section 3: The overall structure is good which includes actors and activities but it needs to harmonize with the taxonomy part. It will be good to include text from the Taxonomy to make it more complete and harmonize.
4. Section 4: Thanks to Felix to start the Section 4.1 on Management. Felix will expands that section to include other management functionalities since managing big data computing cluster is different from traditional client/server practices.

**Technology Roadmap**

* Pw Carey has provided minor input to Section 5 of TR document. The changes were not reviewed on call. Pw is still reviewing Section 7 (any update?)
* David will ping the TR authors for sections that are independent from subgroups and make report for next meeting.

**Online Attendee List**

1. Wo Chang
2. Nancy Grady
3. Sanjay Mishra
4. Rick Jones
5. John Rogers
6. Bruno Kelpsas
7. John Dodd
8. PW Carey
9. William Vorhiess
10. Ashok Malhotra
11. Arnab Roy
12. Bob Marcus
13. Pavithra Kenjige
14. Orit Levin
15. Keith Hare

**Online Chat log**

(1:08 PM) Pw Carey, Complance Partners, LLC: Looked good for a bit....

(1:08 PM) Wo Chang (Host, NIST) disconnected.

(1:08 PM) Pw Carey, Complance Partners, LLC: Take care, sir....

(1:09 PM) Pw Carey, Complance Partners, LLC: Yes... its working...

(1:10 PM) Pw Carey, Complance Partners, LLC: As well you should....

(1:10 PM) Pw Carey, Complance Partners, LLC: Respectfully yours, Pw

(1:15 PM) Bob Marcus joined.

(1:15 PM) Pw Carey, Complance Partners, LLC: Did we forget 'Auditors'....?

(1:15 PM) PavithraKenjige joined.

(1:17 PM) Pw Carey, Complance Partners, LLC: Does this include 'Data Custodians' as well as 'Data Governance'....requirements....as well...?

(1:19 PM) Pw Carey, Complance Partners, LLC: Do you think we can define 'Data Ownership'....or is that too much....?

(1:20 PM) John Dodd: is registration or data registry the same as data availability publication

(1:21 PM) Pw Carey, Complance Partners, LLC: We think it's part of that process but not all by it'self....we think...sorta...

(1:21 PM) John Dodd: Or is there a "consent data form" on primary and secondary uses of the data

(1:22 PM) Pw Carey, Complance Partners, LLC: Good point....we agree...we think it should include down to the 'device layer' for CIA/GRC requirements....Pw

(1:24 PM) Pw Carey, Complance Partners, LLC: Yes...we think you're on the right track regarding.....access rights and responsibilities.....

(1:26 PM) Pw Carey, Complance Partners, LLC: Nancy....is there any chance these meetings could include access to a 'soaphub' where documents can be uploaded for the benefit of the attendees....just a thought....Pw

(1:27 PM) Pw Carey, Complance Partners, LLC: Access Rights....perhaps...

(1:28 PM) Pw Carey, Complance Partners, LLC: Good idea....

(1:28 PM) Pw Carey, Complance Partners, LLC: Access Rights Roles & Responsibilities....et cetera...

(1:30 PM) Pw Carey, Complance Partners, LLC: Would meta data be a separate araea...?

(1:32 PM) Pw Carey, Complance Partners, LLC: Do we need to treat meta data separately from the data itself....?

(1:35 PM) Pw Carey, Complance Partners, LLC: What is it called...Hadoop (what)?

(1:36 PM) Pw Carey, Complance Partners, LLC: Bee Zip 2...yes...?

(1:37 PM) John Rogers, HP: Perhaps we are getting too deep into technical solutioning.

(1:37 PM) John Rogers, HP: Regarding Access Policy Contracts, I believe the two branches should be Policy and Agreements rather than formal and informal

(1:38 PM) Pw Carey, Complance Partners, LLC: But what if the technical solutioning stuff is able to hide critical privacy concerns...yes...?

(1:38 PM) Pw Carey, Complance Partners, LLC: Dear John: that sounds better....

(1:39 PM) Pw Carey, Complance Partners, LLC: Policy and Agreements...

(1:39 PM) Sanjay Mishra(VERIZON) disconnected.

(1:39 PM) Pw Carey, Complance Partners, LLC: Now back to the heady techie stuff....what are your thoughts regarding techno-obfuscation....?

(1:41 PM) Pw Carey, Complance Partners, LLC: Dear John: We're coming from a digital forensics perspective...in this regard....Pw

(1:41 PM) John Rogers, HP: It looks good

(1:41 PM) Sanjay Mishra(VERIZON) joined.

(1:44 PM) Pw Carey, Complance Partners, LLC: Does this Hadoop solution make digital forensics easier or harder....?

(1:46 PM) Pw Carey, Complance Partners, LLC: Can this compressed copy be modified without detection...?

(1:52 PM) Pw Carey, Complance Partners, LLC: Where are the audit trails...?

(1:53 PM) Pw Carey, Complance Partners, LLC: They still are obligated to satisfy Regulatory and Compliance obligations....Pw

(1:53 PM) William Vorhiess, Predictive Modeling, LLC. disconnected.

(1:54 PM) Pw Carey, Complance Partners, LLC: As there are both Internal Auditors as External Auditors....

(1:55 PM) Pw Carey, Complance Partners, LLC: This all revolves around the growth of FRAUD across all industry sectors and jurisdictions.....

(1:56 PM) Pw Carey, Complance Partners, LLC: As associated with Big Data modification and manipulation.....to acheve various objectives, including financial fraud....Pw

(1:59 PM) Pw Carey, Complance Partners, LLC: And lest we forget...the Data Management and Data Archival Processes....yes...?

(2:00 PM) Pw Carey, Complance Partners, LLC: Including, Data Destruction.....(aka: shelf life...)

(2:03 PM) Pw Carey, Complance Partners, LLC: Data Charging Mechanisms.....Pw

(2:03 PM) Pw Carey, Complance Partners, LLC: Machine Learning....w/o human intervention....Supervised and Un-Supervised Machine Learning....yes...?

(2:04 PM) Pw Carey, Complance Partners, LLC: Who is responsible for Un-Supervsed Machine Learning....?

(2:06 PM) Pw Carey, Complance Partners, LLC: In regards to where does the Liability fall....?

(2:08 PM) John Rogers, HP: Hypothesis formulation?

(2:08 PM) Pw Carey, Complance Partners, LLC: Forecasting; good behavior, bad behavior, profitable behavior and a total waste of time and monies....yes...?

(2:08 PM) Orit Levin (Microsoft) joined.

(2:09 PM) Pw Carey, Complance Partners, LLC: Dear John: Can you share an example and/or use case....for us...?

(2:09 PM) Pw Carey, Complance Partners, LLC: Welcome back Orit....

(2:10 PM) John Rogers, HP: In our DARPA programs we explored and tested many data analytic approaches before we found 5 or 6 that provided relevant business solutions. This is different from the "recommender" systems we see in Netflix.

(2:11 PM) John Rogers, HP: The latter Netflix is more indicative of commercial solutions. The government projects tend to be more experimental, as one offs.

(2:11 PM) Pw Carey, Complance Partners, LLC: Dear John: Is there a White Paper or Concept Paper that documents this we could review....? Thanks, Pw

(2:11 PM) John Rogers, HP: Human in the loop is good approach.

(2:12 PM) John Rogers, HP: Exploratory Analytics may be a usefule title in lieu of Human in the Loop

(2:13 PM) Pw Carey, Complance Partners, LLC: Ok...thanks....Pw

(2:14 PM) Pw Carey, Complance Partners, LLC: Enabling Interactive Data Exploration over Big Data | Intel Science ... www.istc-bigdata.org/ index.php/ enabling-interactive-data-exploration-over-big-data/ - View by Ixquick Proxy - Highlight Oct 9, 2013 ... This is an emerging form of data-intensive analytics in which users ask ... They are not designed for human-in-the-loop (interactive) usage and ...

(2:18 PM) Pw Carey, Complance Partners, LLC: Dear Sanjay...our email address is....pwc.pwcarey@gmail.com....

(2:20 PM) Rick Jones (Joseki Group LLC): Gartner recently said that the variety problem accounted for the bulk of their client big data needs.

(2:24 PM) John Rogers, HP: Diferentiate between OLAP and OLTP as these have a major impact on Big Data architecture decisions.

(2:25 PM) John Dodd disconnected.

(2:26 PM) Keith Hare, JCC Consulting, Inc. joined.

(2:29 PM) John Rogers, HP: Data Analytic processes include both Read and Write problems, not just Data Query Processes.

(2:29 PM) Pw Carey, Complance Partners, LLC: Dear Fellow NBD-WG Team Mates....We agree with everything Orit says and votes on, as we must break away for another luncheon function....best wishes and we'll go to work on Section 7, too....Respectfully yours, Pw

(2:29 PM) Sanjay Mishra(VERIZON) disconnected.

(2:30 PM) Sanjay Mishra(Verizon) joined.

(2:38 PM) John Rogers, HP: Wo/Nancy: can we have a copy of the mind map so that we can discuss it with our internal resources at HP. We may be able to help separate some of the Level 2 and 3 activities.

(2:41 PM) Wo (guest): JohnR: if you can send me an email so that I can grant you the access to the mind map.

(2:42 PM) John Rogers, HP: john.w.rogers@hp.com

(2:43 PM) Bob Marcus: Yes. I am online not phone

(2:43 PM) Pw Carey, Complance Partners, LLC disconnected.

(2:44 PM) Nancy Grady (SAIC): I can extract the mmap file and post it (I paid the $6/month to be able to export), or you can ask Wo for permission and work on the mindmeister site

(2:47 PM) Keith Hare, JCC Consulting, Inc.: INCITS -- the InterNational Committee for Information Technology Standards

(2:47 PM) John Rogers, HP: WOuld be happy to participate. Thank you.

(2:47 PM) John Rogers, HP: John

(2:48 PM) John Rogers, HP: Can you text us the address on this box

(2:57 PM) Nancy Grady (SAIC): http://www.mindmeister.com/322462463 but I think Wo needs to grant edit rights, but you may be able to see it anyway

(3:02 PM) Bob Marcus: I'm still online but not on the phone

(3:02 PM) John Rogers, HP: I'll probably need to pay the fee. I'll wait for Wo to respond.

(3:04 PM) Nancy Grady (SAIC): The fee is only to be able to export. You can edit on the site once Wo grants access

(3:04 PM) John Rogers, HP: ok. txs

(3:06 PM) Bob Marcus: No questions at this time

(3:09 PM) Rick Jones (Joseki Group LLC) disconnected.

(3:10 PM) PavithraKenjige disconnected.

(3:15 PM) Arnab Roy (Fujitsu) disconnected.

(3:22 PM) John Rogers, HP: agree

(3:32 PM) John Rogers, HP: Schema on Read is the parlance most frequently used by the No\_SQL market makers

(3:33 PM) Bruno Kelpsas [Microsoft Vendor]: Having to sign-off.

(3:33 PM) Bruno Kelpsas [Microsoft Vendor] disconnected.

(3:50 PM) Bob Marcus disconnected.

(3:51 PM) Keith Hare, JCC Consulting, Inc.: Wo's email is Chang, Wo L (wchang@nist.gov)

(3:51 PM) Nancy Grady (SAIC): nancy.w.grady@saic.com

(3:52 PM) Ashok malhotra (Oracle) disconnected.

(3:52 PM) John Rogers, HP disconnected.

(3:52 PM) Orit Levin (Microsoft) disconnected.

(3:52 PM) Keith Hare, JCC Consulting, Inc. disconnected.