**NIST BD joint subgroup meeting minutes for Dec 23, 2013**

Prepared by: Sanjay Mishra

**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 23, 2013**

**Agenda:**

1. General NBD-PWG status and plans – Wo
   1. RDA Big Data Infrastructure WG - agreed by BDA IG and wait for formal RDA approval
   2. NBD-PWG plans
      1. Finalize V1.0 documents and begin V2.0 early next year
      2. Refresh NBD-PWG subgroups and align collaborations with RDA BDI-WG
   3. Announce ISO/IEC JTC 1 Big Data Study Group with the following meetings:
      1. 1st meeting: March 18 - 21, 2014, San Diego Supercomputer Center, San Diego, US
      2. 2nd meeting: May 13 - 16, 2014, Universiteit van Amsterdam, Amsterdam, Netherlands
      3. 3rd meeting: June 16 - 19, 2014, Beihang University, Beijing, China

***Note****: at each 4-day meeting: first two days for workshop presentations and last two days for SGBD standards work*

* 1. Technical editing - delay due to staff not available yet

1. Arnab and his S&P subgroup to provide an update version of the S&P
2. Review revised text for M0266 RA document – Orit
   1. Text for automated provisioning – David
   2. Text for automated deployment management and recovery – David
   3. Text for Macro level resource management – David
   4. Any further text for System Management – Felix
   5. Text for data management under System Management – David
   6. Anything else needed before technical editing – All
3. Is Definition document ready for technical editing? – Nancy
4. Review and finalize Taxonomy document – Nancy
5. Review and discuss Technology Roadmap document – David
   1. Overall document status – David
   2. More detail discussion on TR Section 7 - Pw

**Notes from the Meeting, Monday, 12/23/13**

**1a.** Wo gave an overview of Research Data Alliance (RDA) Big Data and NIST PWG to collaborate

1. RDA Big Data Analytics Interest Group (BDA-IG) has agreed to form our proposed Big Data Infrastructure Working Group (BDI-WG). Once the BDI-WG is formalized under RDA, BDI-WG reflector will be set up and interested participants can join and work.
2. The working relationship between NBD PWG and BDI-WG are as follow:
   1. NBD PWG will select 5-6 unique Big Data scenarios out from the collected 51 use cases.
   2. RDA BDI-WG will implement these scenarios using NBD PWG Reference Architecture (RA) along with whatever Big Data technologies or solutions might best fit. The end result is to produce best practices implementation guide for how to implement Big Data applications using NBD PWG RA.
   3. NBD PWG will examine and aggregate #b to identify general interfaces between RA five key architecture components. The goal is to identify general RA interface between the RA key components.

**1b.** Current focus is to finalize all NBD PWD V.1.0 working documents to capture all the key technical content. Due to minor glitch of getting technical writers to polish the V1.0 documents, once the V1.0 is finalized, we can put it aside while starting V2.0 starting early January which will last over the next 6 months in 2014.

**1c.** ISO/IEC JTC has announced the BD Study Group to be to the JTC 1 reflector. Three meeting events have planned as listed from today’s meeting agenda. No cost for attendance. INCITS has not yet made a call for participation. Wo will forward once INCITS makes the announcement. Joining the study group is no cost. First two days will be for workshop presentations followed by two days standards work. All submitted papers will be published under NIST Special Publication. High quality with peer-review may potentially be submitted to the ACM conference proceeding. There can possibly be 3 volumes: one from the US meeting, one from Europe and one from Asia, or may be combined into one big volume.

* **Lisa Martinez (question):** Wo, Is there a reason you are aware of for creating a new ISO standard versus ISO20252 is the International Quality Control Standard for Market, Opinion and Social Research, ISO 9001 requires organization to determine, collect and analyze data to ensure the effectiveness of the QMS and whether any continual improvement can be made. ISO 8583 Financial card originated messages — Interchange message specifications is the International Organization for Standardization standard
* **Wo (response):** There are currently multiple groups (over 100 working groups and about 41 subcommittees) within ISO/IEC JTC 1. So before any decision is made WRT BD activity, a study group is spawned with the objective to studying landscape, what is needed, what is the problem, what is the solution, what is the standards gap? The output of the SG will be to, provide specific problems, solutions and identify lack of if any in the current standards. Based on those findings and recommendations, ISO/IEC JTC1 can review those and determine if a WG or a subcommittee needs to be established.
* **Keith Hare, JCC Consulting, Inc. (response):** In addition to Wo's comments, there is not yet a decision to create a new ISO/IEC standard. This is JTC1 Study Group.

**1d.** **Technical writer** funding was requested and approved, however, due to the US Government shutdown, the paperwork still is in progress. Wo to provide additional updates as they become available on possible start date of the technical writer. In the meantime, the NBD PWD V2.0 activity can start in parallel.

**2. Update on Security and Privacy (Presenters: Mark Underwood, Lisa Mangold and Arnab Roy)**

Updates from Mark Underwood (Document Reference: M0274)

1. Explored S&P beyond Velocity, Variety and Volume (domain specific S&P such as Healthcare, M2M etc)
2. The S&P have been improved with better integration of layman and technical concerns. Added additional relevant S&P references (Appendix 1). Added deep dives in S&P in Appendix 2.
3. Remaining tasks include subdocument navigation aids, Taxonomy/RA Integration, Appendix 2 and include additional role/role axes. Role categories (Stakeholders and Users, Ratifiers, Attackers in the (scope section). Role Framework Axes (IT Operator, Governance, Regulation. Compliance officers, Business and Data Consumers)
4. Have not mapped with Taxonomy document yet. (Arnab will also synch up with Orit on alignment with RA)

**Update from Arnab (Reference: M0273)**

Working on the following topics:

1. Taxonomy – Conceptual Axis
   1. Privacy
   2. Provenance
   3. System Health
2. Taxonomy – Operations Axis
   1. Registration, Security Model and Policy enforcement
   2. Identity and Access Management
   3. Data Governance
   4. Visibility and Infra Management
3. Taxonomy of Security and Privacy topics (**Reference Document: M0275**)
   1. Registration, Security Model and Policy Enforcement
   2. Identity and Access Management
   3. Data Governance
   4. Visibility and Infrastructure Management
   5. Risk and Accountability

**Update from Marcia Mangold (Document Reference: NIST Big Data Internal Security Practices.pptx; M0276)**

1. Internal Security Practices with goal to standardize on a security reference architecture. It touches every area (data, people, systems)
2. Deliverables:
   1. Produce a working draft for BD S&P Requirements
   2. Produce a working draft for BD S&P Reference Architecture
3. Scope
   1. Fault tolerance
   2. Difference between BD Security and traditional security concerns
4. Classification of Security and Privacy Topics (classified in 4)
   1. Infrastructure Security
   2. Data Privacy
   3. Data Management
   4. Integrity and Reactive Security
5. Sub categories of Security and Privacy Topics
6. The 5 constructs
   1. Categorize
   2. Select
   3. Implement
   4. Authorize
   5. Monitor
7. GRC and CIA regulatory…
8. Internal Access Control Rules
   1. Overreaching
   2. Access should only be granted
   3. Protection should be based on classification of the data, least privilege and separation of duties that can reduce the risks
9. Access Controls
   1. Most accounts share between different Systems (who is accessing becomes critical)
   2. Trade-off between
10. General Requirements
    1. Role based
11. User and System (Security Controls)
12. Administrative (Security Controls

**Wo (Question):** What is the Difference between Role based versus Access Control?

**Marcia (Response):** Will add distinction and benefits. (will work on it starting in 2014)

**Wo (Question):** What are the time frames the newly additional text will be in place

**Arnab (Response):** One version will be released today. Additional version will be targeted for January.

**Wo (Question):** Can this subgroup help with S&P use case that can be implemented? Is there any way to implement any of the use case (ie health care, retail etc) implement using S&P (role based, access control to manage say DOS)? This will address questions asked by Bob Marcus about implementation approach or issues on S&P Big Data application scenario.

**Arnab (Response):** Is there any example or a blueprint that can be used? Perhaps, this can be started out with Map S&P Taxonomy to RA and map use cases to S&P

**Wo (Question):** What is the plan to map S&P Taxonomy to Mind map to the general taxonomy? Work with Nancy

**Arnab (Response):** Yes, do plan to update.

William Miller (Question): What standard for the metadata? Where are we capturing definition of Meta Data? How much meta data can they access and how it will be used?

Arnab (Response): Can use cryptographic (e.g. timestamp or other attributes validating the origin) and semantics validation.

**Reference Architecture (Orit)**

Orit will incorporate the following topics text sent by David into the RA working draft (M0226):

1. Text for automated provisioning – David
2. Text for automated deployment management and recovery – David
3. Text for Macro level resource management – David
4. Text for data management under System Management – David

Orit will also identify what other text are missing from the M0226 and notify them via email reflector.

**Technology Roadmap (David)**

David will take a closer look at the TR working draft and try to sync up inputs from various subgroups and gather additional inputs from other authors for the rest of the TR sections.

**Action Items:**

1. Arnab will publish S&P today, Dec 23.
2. A second version to be published in January 2014.
3. Big Data S&P scenarios that RDA can implement
   1. How the flow between different sub components
4. Orit will revise M0226 with David’s text
5. Orit will identify any further missing text from M0226
6. David will report the status of the TR at the next meeting
7. Wo will make announcement via email reflector for next NBD PWG telecon

**Next Steps**

1. Finalize RDA, Big Data Infrastructure WG
   1. Identify unique use cases
   2. Workaround the workflow around 5 key components
   3. Technology needed
2. May set up a bi-weekly meetings (jump start sub-groups) and perhaps start a bi-weekly meeting for the next 6-month working period.
3. The focus will be on version 2.0
4. Further JTC1 BD SG telecon meetings (as needed) can be discussed during the first meeting in San Diego

**Online Attendee List**

1. Wo Chang
2. Sanjay Mishra
3. Arnab Roy
4. Geoffrey Fox
5. Orit Levin
6. Keith Hare
7. Bill Mandrick
8. Lisa Martinez
9. Ian G
10. Mark Underwood
11. PW Carey
12. Marcia Mangold
13. David Boyd
14. TS (only initials)
15. William Miller
16. Vivek Navale

**Online Chat log**

(12:46 PM) David Boyd joined.

(12:50 PM) David Boyd disconnected.

(12:50 PM) David Boyd joined.

(12:50 PM) David Boyd disconnected.

(12:51 PM) David Boyd joined.

(12:52 PM) Wo (guest) joined.

(12:57 PM) Marcia Mangold joined.

(1:00 PM) Sanjay Mishra(Verizon) joined.

(1:01 PM) Arnab Roy (Fujitsu) joined.

(1:01 PM) Orit Levin (Microsoft) joined.

(1:02 PM) Orit Levin (Microsoft) disconnected.

(1:02 PM) Keith Hare, JCC Consulting, Inc. joined.

(1:03 PM) Bill Mandrick (Data Tactics) joined.

(1:04 PM) Orit Levin (Microsoft) joined.

(1:04 PM) Geoffrey Fox joined.

(1:05 PM) Lisa Martinez joined.

(1:08 PM) Ian G (in meeting, just watching) joined.

(1:24 PM) Mark Underwood (Krypton Brothers) joined.

(1:26 PM) Lisa Martinez: Wo, Is there a reason you are aware of for creating a new ISO standard versus ISO20252 is the International Quality Control Standard for Market, Opinion and Social Research, ISO 9001 requires organization to determine, collect and analyze data to ensure the effectiveness of the QMS and whether any continual improvement can be made.  ISO 8583 Financial transaction card originated messages — Interchange message specifications is the International Organization for Standardization standard

(1:30 PM) Lisa Martinez: Thank you  your answer was helpful.

(1:31 PM) Keith Hare, JCC Consulting, Inc.: In addition to Wo's comments, there is not yet a decision to create a new ISO/IEC standard. This is JTC1 Study Group.

(1:31 PM) Lisa Martinez: Thank you keith

(1:38 PM) Pw Carey, Compliance Partners, LLC joined.

(1:39 PM) Pw Carey, Compliance Partners, LLC: Sorry we're late....and thank you for the Reminder email....makes life easier on our end.....Respectfully yours, Pw

(1:41 PM) Ts joined.

(1:44 PM) Pw Carey, Compliance Partners, LLC: From our brief research it appears that each of the many Standards Setting Bodies have a unique and separate agenda/criteria that is driving their Standards Process.....Pw

(1:50 PM) Orit Levin (Microsoft): What is the num of the document?

(1:51 PM) Orit Levin (Microsoft): not the presentation, the word doc result?

(1:52 PM) Orit Levin (Microsoft): ah-ha...

(1:52 PM) Orit Levin (Microsoft): Looking forward to reading it ;-)

(1:53 PM) Pw Carey, Compliance Partners, LLC: Glad to see Data Provenance made it onto our Radar screen....Pw

(1:54 PM) Pw Carey, Compliance Partners, LLC: Also, believe we should take a good long look at 'Meta Data' as well...Pw

(1:54 PM) Mark Underwood (Krypton Brothers): definitely... "Veracity" ... Some others call it "Validity" ... No "V" words work perfectly well

(1:54 PM) william miller joined.

(1:56 PM) william miller disconnected.

(1:59 PM) william miller joined.

(1:59 PM) Ts disconnected.

(2:05 PM) william miller: the list states big data for security or is it security for big data?

(2:05 PM) Lisa Martinez: First clue-a person targeting a single person across systems-plus classification

(2:06 PM) Pw Carey, Compliance Partners, LLC: Also, the various states of data, such as; data at rest, in transit, et cetera.....plus the achilles heels associated with data re-useability, and data ACID....as well....

(2:06 PM) william miller: what metadata definition exposes charteristics of the data?

(2:06 PM) Lisa Martinez: We need to come back to you WO-i agree

(2:07 PM) william miller: what standard for the metadata?

(2:07 PM) Pw Carey, Compliance Partners, LLC: Which we're hearing that Big Data is all that great at meeting the 4 ACID Requirements....(aka: 1 of 4, 2 of 4 and/or 3 of 4) but not all 4 of 4.....Please correct us if we're wrong on this.....Pw

(2:11 PM) Arnab Roy (Fujitsu): Folks, Marcia's presentation is doc 276

(2:12 PM) Pw Carey, Compliance Partners, LLC: Ok....Thanks....

(2:17 PM) Pw Carey, Compliance Partners, LLC: Isn't there any way we can incorporate Big Data's capability to analyze, forecast and predict bad behavior, eg fraud.....? Respectfully yours, Pw

(2:21 PM) Mark Underwood (Krypton Brothers): PW It was a use case I had contributed; maybe RDA would choose it. This firm (<http://stealth.ly/>) is using Apache Kafka on the security side

(2:27 PM) Lisa Martinez: Isnn;t the hadoop solution allowing shared system accounts?

(2:30 PM) Pw Carey, Compliance Partners, LLC: Ok...&...thanks...we'll go to that web site....and review...it...Pw

(2:32 PM) Vivek Navale joined.

(2:33 PM) Lisa Martinez: it is very helpful

(2:33 PM) Pw Carey, Compliance Partners, LLC: Thanks we're there...Apache Kafka et al....

(2:36 PM) william miller disconnected.

(2:37 PM) Pw Carey, Compliance Partners, LLC: We'll send them an email...as that's about all we get from their Web site.....

(2:40 PM) william miller joined.

(2:44 PM) Vivek Navale: That means "fixity" has to be placed for both data and meta-data.

(2:46 PM) Mark Underwood (Krypton Brothers): William, if you have use cases, it would be great to study them

(2:48 PM) Pw Carey, Compliance Partners, LLC: When is our next meeting.....and the reminders are quite useful, too....?

(2:56 PM) william miller disconnected.

(2:56 PM) Arnab Roy (Fujitsu): bailing out for 10 mins

(3:03 PM) Pw Carey, Compliance Partners, LLC: Happy Landings......Pw

(3:03 PM) Sanjay Mishra(Verizon) disconnected.

(3:07 PM) Vivek Navale disconnected.

(3:09 PM) Lisa Martinez disconnected.

(3:10 PM) Vivek Navale joined.

(3:15 PM) Pw Carey, Compliance Partners, LLC: Yes...we're filling in the gaps for Section 7....text/content...yes we are...Pw

(3:16 PM) Pw Carey, Compliance Partners, LLC: Our audio is breaking up at this end....sorry...Pw

(3:17 PM) Pw Carey, Compliance Partners, LLC: We sent you an email Sunday, which may not have reached you, we'll re-send it.....lastly, hope you have a nice and relaxing Merry Christmas and our best wishes, too....

(3:19 PM) David Boyd disconnected.

(3:19 PM) Mark Underwood (Krypton Brothers) disconnected.

(3:19 PM) Orit Levin (Microsoft) disconnected.

(3:19 PM) Arnab Roy (Fujitsu) disconnected.

(3:19 PM) Keith Hare, JCC Consulting, Inc. disconnected.

(3:20 PM) Pw Carey, Compliance Partners, LLC: Take care...see you all in the New Year.....Pw

(3:20 PM) Wo Chang (Host, NIST): thank you all for all your hard work; Merry Christmas and  a Happy New Year; rest up and get busy again starting early next year.