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

**NBD-PWD-2015/M0436**

**Source: NBD-PWG**

**Status: Draft**

**Title: Web chat and note from Meeting of Sept. 1, 2015**

**Thanks to Russell Reinsch for taking note!**

*Chat Log C:\underwood\gov\NIST\Big Data\meeting-20150901\ChatLog 2015\_09\_01 14\_37.rtf*

*Gregory Weidman (Private): 1:01 PM: I'm glad to be here. I have a hard stop for me at 1:40. Sorry i won't be here for the whole thing.*

*Mark Underwood (to Gregory Weidman): 1:01 PM: Gregory - Roger, thanks.*

*Russell Reinsch (to Everyone): 1:04 PM: Sorry I had my speaker off and I didnt notice the app had opened and I had background noise going*

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:13 PM: We'll put ourselves on mutae......for the benefit of all......*

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:14 PM: Who is funding the CSRA....?*

*gretchen gottlich (to Everyone): 1:14 PM: agreed*

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:16 PM: ensia European Union Agency for Network Security.....we'll have to go reach out for the .pdf taxonomy effort......*

**Mark starts the meeting discussion on controlled vocabulary and taxonomy for the NIST S&P and shows a few sources for where we might find existing work that we can use all of or part of to speed up the development of the NIST S&P taxonomy. He covers HL7s ontology and Enisa ontology and a few others. The idea that maybe we can combine some of them into something useful.**

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:17 PM: Dear Mark: Good effort.... will these be available for us to download and review.....good effort, deux....*

**Jacob mentions that he does not recommend attempting a taxonomy.**

**Gregory is concerned about getting outside of the realm of big data.**

**Tim cites two examples of ontology failures. In the first, the ontology caused total failure of the group because people couldn’t agree on terms. He offers the solution: that we focus on federation, authorization, auditing; as an ontology will always end up unresolvable.**

**Mark basically explains that we need something; and reiterates that he was not suggesting that we attempt a large or drawn out taxonomy project; just suggesting that we might adopt some existing work from other sources and mash them up.**

**Arnab begins presentation of mo343 and provides background for people on the call and suggests that the BDPWG does have a need to classify things… Mo343 has a slide showing 2 aspects of BD S&P: scaling and mixing.**

**The scaling aspect involves the volume (and one of the other Vs) and mixing involves two other Vs. Mixing includes access and he makes some discussion into privacy etc.**

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:23 PM: Should we be looking at Deep Learning....and the major division of 'data'.....Recorded data from the physical world.....and 'data' created by you all.....content creation.....?*

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:23 PM: This would be the major division of Big Data into two buckets.....no?*

**Arnab con’t: NIST S&P Conceptual topics: privacy, system health, provenance. Continued background on our group’s attempt to map the S&P volume to the RA volume and how the two were being developed in parallel / congruently. The Operational aspect of the S&P taxonomy is more related to the RA due to the RA’s theme more about technologies and deployment.**

**Arnab now discussing mo387 and two taxonomies. Two approaches to \_. One approach = top down. Start with the data Requirements, which are “human.” These are on the left hand side of the diagram and yellow. They relate to the Concepts which are intermediately technical (vs. human) and all of the concepts are managed by the System Orchestrator in the top center of the diagram. Lines in this Visio diagram show connections between requirements, concepts, other entities below.**

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:25 PM: With follow along sub classifications/iterations.....*

*sharafudheen (to Everyone): 1:27 PM: I am not able to get audio... in the system ..*

*sharafudheen (to Everyone): 1:27 PM: anyone else experience this issue?*

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:28 PM: Please Note: FinTech via Big Data will take us into the commercialization of Big Data.....*

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:28 PM: Our sound is good via Skype....dial up*

*Mark Underwood (to Everyone): 1:29 PM: Russell- can u take notes for us?*

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:29 PM: If you're running in Windows....you may want to re-boot and try again.....*

*Russell Reinsch (to Everyone): 1:29 PM: Sure*

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:30 PM: Dear Mark: We'll send you a copy of the Chat...chat, as our setup generates a copy of each session's Chat Line......not sure why...*

*Jacob Dilles (to Everyone): 1:34 PM: +1 Tim*

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:35 PM: Dear Tim: Interesting.....but we're not the final word in this arena....just a gathering of folks with a deep interest in creating something useful in the real world.....this group is famous for it's ego-less culture which comes from the top.....down.....*

*Tim Zimmerlin (to Everyone): 1:38 PM: Federation: IPsec, Identity provider, Identity manager, Certificate authority, RA, etc.*

**Jacob is concerned about usage of the term Concepts and clarifies that this term has very diff meaning in other contexts / groups. He recommends changing Requirements to Business Requirements and changing Concepts to Technical Requirements.**

**Arnab is not opposed to those changes.**

**Mark queries Jacob regarding ISO docs, that are not trivial to wade through, on how to proceed with checking out ISO terms. Jacob suggests 704 might be the best doc to look at.**

**Arnab notes that there are two parts to the S&P effort; the S&P taxonomy which is text in Word, sort of a flattened partner to the Visio diagram which is able to show the connections between entities. The S&P taxonomy is now displayed onscreen and Arnab clarifies that “Concept” is an abstraction “at the level of a system orchestrator.”**

**Tim suggests that we could possibly use the terms “services” or micro services instead of “concepts.” Also that we could include a clarification sentence at the beginning of \_ to clarify for readers that we are referring to concepts here as abstractions and not whatever any other usage of the term might imply to different audiences.**

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:38 PM: Welcome back sharafudheen....Pw's advice must have been spot on......what a genisus (sp)....no?*

*Gregory Weidman (to Everyone): 1:39 PM: Thanks, Mark, for a great discussion! As I mentioned earlier, I need to step away for another meeting.*

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:39 PM: Dear Folks: Can we get a copy of this Unified Security and Privacy Taxonomy Sep. 1, 2015....please....Thanks.....*

*Arnab Roy (to Everyone): 1:40 PM: 343 on the NIST website*

*Arnab Roy (to Everyone): 1:41 PM: http://bigdatawg.nist.gov/\_uploadfiles/M0343\_v1\_8647080366.pdf*

*Tim Zimmerlin (to Everyone): 1:48 PM: The first SnP framework could identify "what is being protected" and "what are the threats" and "what are the SnP methods used".*

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:49 PM: Dear Arnab: Thanks....*

*gretchen gottlich (to Everyone): 1:49 PM: thank you well done to the team*

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:50 PM: Dear Arnab: Follow-up....we believe we should shift our attention to Deep Learning......within a Big Data Eco-system......*

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:50 PM: & it's impact upon Big Data.....*

*Tim Zimmerlin (to Everyone): 1:51 PM: Pw, "deep learning" is not a well defined concept…just like "big data".*

*sharafudheen (to Everyone): 1:52 PM: I have send a academic paper on the Big Data security framework...*

*sharafudheen (to Everyone): 1:52 PM: I am suggesting a consolidated framework in this paper.. could you please look into it?*

*sharafudheen (to Everyone): 1:53 PM: something similar to the one arnab is showing in visio drawing*

*Mark Underwood (to Everyone): 1:53 PM: sharafudheen - you have a URL?*

*sharafudheen (to Everyone): 1:53 PM: can someone acknowledge the receipt of the document in your email?*

**Mark notes that the S&P taxon onsceeen does not give the appearance of being about big data / rather it could be about computing in general and there is a bit of discussion from ~4 participants on reducing work so that it is limited to big data; and commentary that without context of what is not big it is difficult to actually represent big. Mark: there is a need to cover general issues in order to descrie big data issues; it sets the scene / of what is diff about big data. Also that we started with the Vs, and if we take out the background then we have only computer science…. Something.**

**Arnab clarifies that the requirements in each of these sub categories are different from general computing.**

*Tim Zimmerlin (to Everyone): 1:53 PM: In my current work, experts are discussing three big data orchestration SnP frameworks: ssh, REST, and SOA.*

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:54 PM: Dear Tim: Correct....not well defined but growing expoentially......ten to the power of ten.....and will become a game changer within the world of Big Data.....in our professional and personal opinion.....(aka: can't we provide clarity in this area of Big Data and Deep Learning....?*

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:55 PM: Dear Tim: One area such as FinTech....as one area....of interest and growth....*

*Tim Zimmerlin (to Everyone): 1:56 PM: Big data experts have debated the strengths, weaknesses, and roles of each. Today, REST has the largest following and it is growing. Tomorrow, ssh will be too labor intensive for 10's or 100's of thousands of cores.*

*gretchen gottlich (to Everyone): 1:56 PM: agreed on new labels*

*Tim Zimmerlin (to Everyone): 1:57 PM: Pw, the technical art is moving too quickly to standardize in the nearterm.*

*gretchen gottlich (to Everyone): 1:58 PM: <hehe> mark*

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:58 PM: We agree with gretchen.....and Tim: can't we play a bit of catch-up in this area if it is truly going to play in the real world of business.....?*

*Mark Underwood (to Everyone): 1:59 PM: Russell - for the notes: ISO 704; these are docs for how to proceed with terms in that world*

*Tim Zimmerlin (to Everyone): 1:59 PM: Pw, not really…not even close.*

*Pw Carey, Compliance Partners, LLC (to Everyone): 1:59 PM: Dear Tim: Bummer.....*

*gretchen gottlich (to Everyone): 2:00 PM: ty*

**… the conversation gets to Jacob discussing how something starts out as an original work; experts get involved; and the work evolves to a point where the authors no longer need to justify their product / documentation.**

**Jacob mentions some disagreement amongst ISO on whether Orchestrator is even a role.**

**Arnab states that we can extrapolate what is new about big data in the text of our documents.**

**Jacob essentially agrees and suggests that authors can reference a source for the pre-existing knowledge (sources in other docs etc.) and carry on with descriptions about what is new about big data.**

**Mark queries Arnab? About challenges in version 2 such as bringing the Vs back in, or how we might deal with the loss of the “Orchestrator” role.**

**Jacob reiterates for a second time that we should not be trying to mirror ISO. We should frame things correctly for NIST and there would be less / no problem.**

**Arnab clarifies that an architect would be the one who decides how a project progresses from requirements to technologies.**

**Jacob agrees and notes that he personally thinks that the Orchestrator must be included. He clarifies that ISO is typically not guidance on implementation (vs. NIST which is); and also that the ISO RA is really not an RA, for the same reason.**

**See Tim’s comment in the chat at 2:22 (replying to PW’s 2:19) and Jacob’s 2:24 reply.**

*sharafudheen (Private): 2:05 PM: Hi Mark.. kindly check ur mail and if possible please share the attachment to the team.*

*sharafudheen (Private): 2:06 PM: My Mail: "Dear Mark,*

*This is Sharafudheen from Big Data working group.*

*Kindly see the attached academic paper that I completed as part of my PhD work.*

*Kindly share this document with the group, as it is waiting for the approval.*

*If you can do it and send a message to the team during the meeting, it will be great..."*

*sharafudheen (Private): 2:07 PM: I contain some relavant information related to S&P group... I have complied the archetecture (basic one).. in a different way... it is just a start....*

*Mark Underwood (to sharafudheen): 2:09 PM: Do you want to present during the meeting, or have us distribute by email after the meeting?*

*sharafudheen (Private): 2:10 PM: I could present it, if you give me a chance*

*Pw Carey, Compliance Partners, LLC (to Everyone): 2:10 PM: Dear Folks: Where is Archival Treatment...along the Opt Out/Opt In criteria....and Individual ...Anomitization .... and what takes this origiinal work into the realm of Big Data is the application multiple interations and filters resulting in a final physical product (aka: analytical in nature similar to data nurons.....) no?*

*gretchen gottlich (to Everyone): 2:11 PM: silly idea here can put this gnarly issues as a discussion in the appendix?*

*Pw Carey, Compliance Partners, LLC (to Everyone): 2:11 PM: Dear Gretchen: Useful idea.....dare we say....a good idea for discussion.....*

*gretchen gottlich (to Everyone): 2:12 PM: or maybe as as a preface? introduction?*

*sharafudheen (Private): 2:12 PM: I got the "best paper award " during the academic conference, btw....*

*Pw Carey, Compliance Partners, LLC (to Everyone): 2:16 PM: Either seems agreeable to my entire R&D Team....*

*gretchen gottlich (to Everyone): 2:16 PM: sounds like maybe multiple "papers" citing eachother...*

*Tim Zimmerlin (to Everyone): 2:18 PM: Amazon AWS sends about 2 million SnP event msgs per day to AWS tenants. In the future, 1 millions of AWS tenants should be able to send SnP event msgs to AWS orchestrators as warnings.*

*Pw Carey, Compliance Partners, LLC (to Everyone): 2:19 PM: Can't we give it our best effort from a NIST perspective...then hand it over the transum to the folks hiding behind the ISO door....no?*

*Pw Carey, Compliance Partners, LLC (to Everyone): 2:20 PM: Dear Tim: Any idea how many of these are false positives....?*

*Pw Carey, Compliance Partners, LLC (to Everyone): 2:20 PM: Perhaps not.....too hard to pin down....*

*Tim Zimmerlin (to Everyone): 2:21 PM: Could the SnP subgroup develop an abstract finite state machine model. E.g., user logs in, IdP authenticates, user requests SnP services from remote system, remote IdM checks with that user's IdP, etc.*

*gretchen gottlich (to Everyone): 2:21 PM: oh certainly!*

*Tim Zimmerlin (to Everyone): 2:22 PM: Pw, poor strategy. The last editor wins the game.*

*Jacob Dilles (to Everyone): 2:24 PM: Tim - careful with 'win' ... ISO is not law, it is to encourage colaboration internationally. They don't have to be identical*

*Tim Zimmerlin (to Everyone): 2:25 PM: Jacob, msg understood.*

*Pw Carey, Compliance Partners, LLC (to Everyone): 2:25 PM: But does this mean it's still not ok to shoot the Editor....?*

*gretchen gottlich (to Everyone): 2:26 PM: wonderful thank you for your graciousness :)*

*gretchen gottlich (to Everyone): 2:26 PM: good question!*

*Pw Carey, Compliance Partners, LLC (to Everyone): 2:27 PM: Looks good....definitely would like to review....sharafudheens efforts.....thanks...*

*gretchen gottlich (to Everyone): 2:28 PM: http://hortonworks.com/hdp/security/*

*gretchen gottlich (to Everyone): 2:28 PM: whitepaper available*

*Mark Underwood (to Everyone): 2:28 PM: Sharafudheen is recommending a look at SAP Hana and Hortonworks sec frameworks* **Correction here: it was somewhat difficult to hear Sharaf; he clarified that he is not recommending Hana…**

*Keith Hare (to Everyone): 2:29 PM: There is a whitepaper "Solving Security Hadoop" on the web page* [*http://hortonworks.com/hdp/security/*](http://hortonworks.com/hdp/security/)

**Sharaf takes us through some pages in his academic paper on big data security which seems to include several references to NIST. Sharif has coined the term or introducing a new role: Security Orchestrator. Asked if he encountered any reference to “fabric” in his research he responds no. Something about Hortonworks; and similarities of some components; and that Horton has one doc online that might be a good resource.**

**Several people on the call give positive feedback to Sharaf and everyone is looking forward to his involvement.**

**Jason comments that SSO is very important.**

**Tim talks a bit about Active Directory and related things.**

*Pw Carey, Compliance Partners, LLC (to Everyone): 2:30 PM: SAP and Security has been a data goblin for quite some time within the Big Data / HANA arena....*

*gretchen gottlich (to Everyone): 2:30 PM: SAP Hana is complex....and not as finished as Hortonworks*

*Mark Underwood (to Everyone): 2:31 PM: Thx Gretchen*

*Mark Underwood (to Everyone): 2:32 PM: Keith, thx*

*gretchen gottlich (to Everyone): 2:32 PM: was working with T-Mobile and SAP Hana; evolving is a good word...*

*Bhumip [ZTE] (to Everyone): 2:35 PM: IETF cross-domain ID use-case draft is at http://datatracker.ietf.org/doc/draft-ietf-scim-use-cases/ .. Thanks*

*Pw Carey, Compliance Partners, LLC (to Everyone): 2:36 PM: Dear Sharafudheen: Good luck....*

*gretchen gottlich (to Everyone): 2:37 PM: https://play.google.com/books/reader?id=LklKCgAAQBAJ&printsec=frontcover&output=reader&hl=en&pg=GBS.ZZ0*

*gretchen gottlich (to Everyone): 2:37 PM: data as a service chapter on big data security.*

*sharafudheen (Private): 2:37 PM: you can call me "SHARAF"*

*Russell Reinsch (to Everyone): 2:38 PM: Miller time*

*Pw Carey, Compliance Partners, LLC (to Everyone): 2:38 PM: Good meeting......*

*sharafudheen (to Everyone): 2:38 PM: you can call me "SHARAF"... thanks mark for giving an opportunity to present*

*gretchen gottlich (to Everyone): 2:38 PM: thank you! meeting very well chaired and valuable!*