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

**NBD-PWD-2015/M0447**

**Source: NBD-PWG**

**Status: Draft**

**Title: Web chat from Meeting of Sept 29, 2015**

**Chat Log D:\\_wo\1DMG\2015\\_BigDataWG\Docs\ChatLog 2015\_09\_29 15\_03.rtf**

**Russell (Private)**: 12:59 PM: Wo I installed the gotomeeting app on my machine but it seems like IM still signing in through the web channel.

**Russell (Private)**: 1:01 PM: Not sure why my newer machines are not getting it setup right / like my older machine did so easily.

**Wo Chang (to Russell)**: 1:02 PM: Yes, looks like you're still on the web. Let's trobueshoot this after the telecon. Thanks.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:02 PM: Dear Sir: Thank you for the Use Case follow-up....Yes please REMOVE CONFIDENTIAL.....didn't even think of that...just a formality we're use to....and thanks again....

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:02 PM: Yes we can and we'll stay on Mutae.....

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:04 PM: Also, we have the ChatLog for our Conf. Calls since June...which we'll send to Mr. Chang for his approval......June thru todays Conf. Call....

**sharafudheen (to Everyone)**: 1:10 PM: I am not able to hear... there is some issue.. tried with two devices... is everyone else OK?

**Nancy Grady (to Everyone)**: 1:11 PM: The sound is coming through the goto meeting site for me.

**sharafudheen (to Everyone)**: 1:11 PM: OK.. it looks some mis-configuration from my end. I will try to figure out

**Ian Gorton (to Everyone)**: 1:11 PM: i can hear Wo

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:16 PM: Data Sets and Data Analysis....The data will be available to run....using the same platform...for comparing notes....we want to make sure the environment is suitable for the 'Use Cases'......

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:19 PM: Handing over to Ian....for the types of students he has.....Ian can you speak....yep...we can hear you....We're working on the Health Care Use Care.....working from Seattle...Summary Health Care Medicare Providers....amount of patients they treat and the costs of reimbursements.....they make......understanding the entities....created 20 or so Analytics they can be based on....eg. number of health care providers in Wash. Stt. to do colon exams...et...

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:20 PM: Looking for outliers....Drs who do back surgery but conducting eye surgery.....as an example an Outlier....

**David W. Boyd (to Everyone)**: 1:20 PM: This is going to make some health care effort and data analytics interesting going forward:

**David W. Boyd (to Everyone)**: 1:20 PM: Melinda Beck reports in the WSJ that doctors, hospitals and insurers are bracing for possible disruptions on October 1 when the U.S. health-care system switches to ICD-10, a massive new set of codes for describing illnesses and injuries that expands the way ailments are described from 14,000 to 70,000. Hospitals and physician practices have spent billions of dollars on training programs, boot camps, apps, flashcards and practice drills to prepare for the conversion, which has been postponed three times since the original date in 2011. With the move to ICD-10, the one code for suturing an artery will become 195 codes, designating every single artery, among other variables, according to OptumInsight, a unit of UnitedHealth Group Inc. A single code for a badly healed fracture could now translate to 2,595 different codes, the firm calculates. Each signals information including what bone was broken, as well as which side of the body it was on.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:21 PM: Probably going to do some Spark to do some of the Machine Learning......using a Web based host using Google Maps or something along those lines......Thanks Ian.....

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:21 PM: That would be Ian Gorton......of course....

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:22 PM: How does the Orchestratrator talk with the data provider......incorporating the RA Reference Architecture and the dataflow.....will need to talk with Jeffrey off-line......

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:23 PM: Looking for a common platform so we can all work together......Ian: Some of the Use Cases may not map to the RA Reference Architecutre....it's hard to see what the Orchestrator is doing ....

**Mark Underwood (to Everyone)**: 1:23 PM: So there is no significant ETL here?

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:24 PM: Dear Mark: Is that correct.....there is no Extract.....Tramsform/Transmit......Load....?

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:27 PM: This will not be tracking individual patients....there will be nothing about individual patients.....is there any risk attached to patient security and privacy.....patient security and privacy issues have all been considered and addressed.....Yes.....No.....? or is all the patient data simply an aggregate collection of patient data....?

**Mark Underwood (to Everyone)**: 1:27 PM: The UCSD Supercomputer center also hosts this data. Not sure they're doing with it

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:28 PM: Dear Ian: If we bounce your aggregate patient zip code data against Facebook.....et al...there is no opportunity for patient identity being uncovered....yes....?

**Mark Underwood (to Everyone)**: 1:29 PM: No intention to be accusatory here. )

**Mark Underwood (to Everyone)**: 1:30 PM: My point was just to understand the use case more deeply

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:30 PM: Dear Mark: Thanks....

**Ian Gorton (to Everyone)**: 1:30 PM: PW - most old people don't have FB accounts ;)

**Ian Gorton (to Everyone)**: 1:32 PM: and even if they did, all you have is teh providrr zip. This is not necessarily the same as the patients. OK - you can cluster/guess, but in most places thee's going to be a lot of candidate pataients

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:33 PM: Dear Ian:....Does this mean your data pool demographics have some age categories......? We're looking for a sneaky back-door into identifying individuals within your data sets.....along the lines of what the bad guys would do.....Thanks....

**Jacob Dilles (to Everyone)**: 1:33 PM: Wo: I don't think you can guarantee there is no complimentary data set... That's why we call BD a "paradigm shift"

**Ian Gorton (to Everyone)**: 1:34 PM: PW - all we know is that they are medicare patients

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:35 PM: Dear Ian: Is that the same Medicare Database that was hacked......? Just wondering....Don't get discouraged.....sounds like a good Use Case.....regardless...Pw

**Ian Gorton (to Everyone)**: 1:35 PM: to be precise - all we know is that a provider treated N medicare patients

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:36 PM: Dear Ian: There is not an impuning bone in our bodies....(aka: the bodies within this WG....)

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:38 PM: Hear hear......

**Ann Racuya-Robbins (to Everyone)**: 1:39 PM: I agree with Ian

**Ann Racuya-Robbins (to Everyone)**: 1:41 PM: If privacy means anonymity I agree with Ian. Howewver if anonymity is not the goal but rather fairness this problem can be rather easily solved

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:41 PM: Dear Ian: Correct....especially once your data is passed around to 3rd parties with or without reasonable controls for accountability.....However would a 'block chain' strategy including encryption....alleviate many of the privay security concerns.....perhaps....?

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:41 PM: We meant privacy...rather than privay....

**Mark Underwood (to Everyone)**: 1:42 PM: At the least, there are governance issues that we are trying to address / Risk Management. and responsbility.

**David W. Boyd (to Everyone)**: 1:42 PM: On the Orchestrator discussion. More and more I am seeing the Orchestrator role as a none-automated and mostly human executed role. As I look at the activities and functional components I came up with for that role it seems to me it is more a requirements/policy role.

**Frank Farance (to Everyone)**: 1:43 PM: Sorry, need to sign off, good discussion!

**Mark Underwood (to Everyone)**: 1:44 PM: Dave - Re Orchestrator - But workflow can be automated as well as interlaced with business processes

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:44 PM: Frank Farance needs to sign off...."Good discussion...."

**Mark Underwood (to Everyone)**: 1:44 PM: Not sold on that, Ian

**Mark Underwood (to Everyone)**: 1:45 PM: If a patient is reidentified from this data, CMS will get a phone call, and the company doing the Big Data Variety join might hear from the FTC

**Ann Racuya-Robbins (to Everyone)**: 1:45 PM: Agree with you Mark. What does legitimately get to the data mean?

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:46 PM: Dear Jacob: That sounds correct to us......as in a Block Chain strategy.....where any changes to the previous entered data.....negates their existence......without any access back to that data.....

**Jacob Dilles (to Everyone)**: 1:47 PM: I'm willing to go offline to discuss further...

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:48 PM: Dear Jacob: Here is our email address for off-line discussion of same.....pwc.pwcarey@gmail.com

**Ian Gorton (to Everyone)**: 1:48 PM: legitimately get to data means use an interface to a system (UI/programmatic) to obtain some (subset of?) the data. It's tricvial - I legitimately got the medicare data using http

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:50 PM: Dear Ian: Looking forward to reviewing your Use Case....sounds quite interesting to us.....Good luck....

**sharafudheen audio (to Everyone)**: 1:51 PM: all the best Ian for ur use case...

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:51 PM: Dear David: Good post....re: Orchestrator....

**Ian Gorton (to Everyone)**: 1:52 PM: so the orchestrator is not a 'run time' thing? Not the interpretation I had to be honest ..

**sharafudheen audio (to Everyone)**: 1:52 PM: i also have research students who can collaborate with u .. if u will.. let us talk offline :sarafudheen@yahoo.com

**Ian Gorton (to Everyone)**: 1:52 PM: cool - email me ... i.gorton@neu.edu

**Ann Racuya-Robbins (to Everyone)**: 1:53 PM: Wo do you mean something like a policy engine?

**Mark Underwood (to Everyone)**: 1:53 PM: +1 for Wo's comment: Seems misleading to characterize workflow as human-in-loop only

**Ian Gorton (to Everyone)**: 1:53 PM: my interpretation was along th elines Wo just described

**Bob Marcus (to Everyone)**: 1:53 PM: Where is an automated workflow for Big Data analytics in the Reference Archotecture?

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:54 PM: Dear Bob: Don't recall there was one.....

**Mark Underwood (to Everyone)**: 1:55 PM: Workflow in the Spotify big data system decides what will be recommended on the user's web page after they play a tune

**Ann Racuya-Robbins (to Everyone)**: 1:55 PM: Agreed Nancy. Still how is uncertainty addressed and described?

**Mark Underwood (to Everyone)**: 1:56 PM: +1 Ian

**Ann Racuya-Robbins (to Everyone)**: 1:56 PM: I think that is right Ian.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:56 PM: Yep.....

**Bob Marcus (to Everyone)**: 1:57 PM: The issue is the difference between run-time and design and configuration phases. This has to be clarified in the RA.

**Ann Racuya-Robbins (to Everyone)**: 1:58 PM: Maybe we need something more than RA!

**Mark Underwood (to Everyone)**: 1:58 PM: CMMI-DEV, a quasi standard, deals with design-side processes

**David W. Boyd (to Everyone)**: 1:58 PM: Bob +1

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:59 PM: We agree....we should clarify....slash qualify...the Orchestrator construct.....what it does and what it doesn't do......lets keep it simple.....

**Ian Gorton (to Everyone)**: 2:00 PM: just for fun ... quabase.sei.cmu.edu

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:01 PM: Dear Ian: Is this a web site....http://www.quabase.sei.cmu.edu......?

**Ian Gorton (to Everyone)**: 2:01 PM: yep

**Ian Gorton (to Everyone)**: 2:01 PM: if its a person/team, to me its a role in a reference architecture

**Bob Marcus (to Everyone)**: 2:01 PM: In the RA diagram, the red SW flows are usually done at configuration time while the blue data flows are usually at run-time.

**Ian Gorton (to Everyone)**: 2:01 PM: I am playing a role in teh project with my students

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:02 PM: Ok...thanks..we'll go there now.....

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:03 PM: Here yah ......https://quabase.sei.cmu.edu/mediawiki/index.php/Main\_Page

**sharafudheen audio (to Everyone)**: 2:03 PM: all the best Ian!!!

**sharafudheen audio (to Everyone)**: 2:05 PM: we have some use cases on "big data for security" from our organization. the scenario is to bring the system logs to Splunk system and analyze the data for potential loopholes and security breaches. do u think this is a valid use case to be considered?

**sharafudheen audio (to Everyone)**: 2:06 PM: the data set is simply unix server log files.

**Mark Underwood (to Everyone)**: 2:07 PM: Splunk use cases are often relevant

**Jacob Dilles (to Everyone)**: 2:08 PM: This is a potential use case but be careful not to get recusive

**sharafudheen audio (to Everyone)**: 2:08 PM: u can mute all

**Mark Underwood (to Everyone)**: 2:08 PM: SharaFudheen - especially Splunk with log files for net defense or forensics

**sharafudheen audio (to Everyone)**: 2:11 PM: i mean to look at it from our archetecture perspective

**Wo Chang (to Russell)**: 2:13 PM: that's great sharafudheen!

**Wo Chang (to Everyone)**: 2:14 PM: that's great sharafudheen!

**Russell (to Everyone)**: 2:17 PM: Wo - Ive emailed you a 6b draft

**sharafudheen audio (Private)**: 2:18 PM: thanks Wo... i will try to bring some scrambled sample data for u...

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:20 PM: Dear Dave: Do you breakout your descriptions based upon industry regulatory requirements, et al.....?

**David W. Boyd (to Everyone)**: 2:21 PM: PW - not really. Those would be valid stakeholder concerns and the regulators would be stakeholders.

**Mark Underwood (to Everyone)**: 2:22 PM: Dave / Wo - I dunno that WG9 is the right venue. The NBDWG v2 seems like a better place to hammer out an approach, esp for orchestration. Eg the privacy issues are really hard in the int'l setting, whereas it's maybe the KEY (sic) issue in the US for Big Data

**Ann Racuya-Robbins (to Everyone)**: 2:22 PM: Wo where can I find David Boyd's docuoment?

**Keith Hare (Private)**: 2:22 PM: Have to drop off. Later.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:23 PM: Dear Dave: But the Regulators are not a party to the design and build phase....my understanding of the world is that the Regulators come into play after everything is built......Please correct me on this.....?

**sharafudheen audio (Private)**: 2:24 PM: if u were looking for my name, call me Sharaf

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:25 PM: Dear Dave: Consider the plight of the Auditor....their requirements and considerations are not included in the design and build phase .... they're treated as an afterthought......No?

**Mark Underwood (to Everyone)**: 2:26 PM: Dave - You mentioned ActiveMQ et al. - I was going to bring up messaging as well. And I remind myself that agent-based approaches in Big Data make the notion of roles more elastic between human, hybrid and purely machine

**David W. Boyd (to Everyone)**: 2:27 PM: end to define requirements. While you may not deal with them in design in all cases, in some cases you do. For example, I work on classified systems - we are regulated by an accreditation authority, who does review every aspect of the design for security, I know we produce a lot of documentation for them.

**David W. Boyd (to Everyone)**: 2:28 PM: Lost the first part of the last message but the start said regulators often define requirements - often related to auditing.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:28 PM: Dear Dave: Ok...thanks.....good to know....this does occur....

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:29 PM: Pw says.....yes, yes, yes....lets defer....also, our microphone doesn't work for some reason which we hope to fix before our next meeting....

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:31 PM: Our Use Case is Twitter Real-time Analysis within a FinTech environment....(Financial Technology)...

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:33 PM: Yes.....of course....Mr. Chang is going to post it to the NBD web site.....once he removes the word 'CONFIDENTIAL' from the document.....

**Ann Racuya-Robbins (to Everyone)**: 2:38 PM: I would like to work on it and appreciate your suggestions Nancy

**Ann Racuya-Robbins (to Everyone)**: 2:39 PM: What was the rationale for making these definitions recursive? I can ask you offline as well Nancy.

**Russell (to Everyone)**: 2:41 PM: A mosaic plot of associations

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:42 PM: Dear Russell: A very nice turn of phrase......Thanks....

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:43 PM: Dear Nancy: Do you still have Ian's site up and can you share it here.....Thanks...?

**Mark Underwood (to Everyone)**: 2:44 PM: For discussion- I feel governance belongs in Vol 4. There will be debate around this, but it's too hairy to crosswalk between the two vols in our current mapping of the v2 work

**Russell (to Everyone)**: 2:44 PM: Nancy: I have started a pattern table; its posted on the PWG site

**Nancy Grady (to Everyone)**: 2:45 PM: @Russell, what doc number?

**Russell (to Everyone)**: 2:45 PM: Let me look

**Nancy Grady (to Everyone)**: 2:46 PM: is it 434?

**Russell (to Everyone)**: 2:46 PM: yes 0434

**Ann Racuya-Robbins (to Everyone)**: 2:47 PM: Thanks

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:49 PM: This document has not been released to the Public yet........for Comments......it has to go through the NIST Publication process.....

**Nancy Grady (to Everyone)**: 2:50 PM: @PW, Ian's site is the one posted earlier https://quabase.sei.cmu.edu/mediawiki/index.php/Main\_Page

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:54 PM: Dear Nancy: Ok.....thanks.....

**Ann Racuya-Robbins (to Everyone)**: 2:55 PM: Mark I would like to understand why agents need to be involved?

**Ann Racuya-Robbins (to Everyone)**: 2:56 PM: For privacy?

**Mark Underwood (to Everyone)**: 2:57 PM: Software agents become proxies for what used to be roles inside a firm - e.g., people used to make credit decisions, but now it's large automated to FICO

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:58 PM: Dear Mark: Not to be argumentative.....according to my spies in the EU......this attitude is changing......as a Class Action Law Suit is progressing against Google within the EU Court of Justice.....and it doesn't look promising for Googles bottom line.....just a thought.....

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 3:00 PM: So....a Role can be both an individual as well as a Job Title.....No?

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 3:01 PM: Nice meeting.......and Drive Safely....

**Ann Racuya-Robbins (to Everyone)**: 3:02 PM: Thank you all.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 3:02 PM: Dear Mark: Thanks for the Heads Up...

**Nancy Grady (to Everyone)**: 3:03 PM: nothing here

**Alicia Zuniga-Alvarado (to Everyone)**: 3:03 PM: Thank you

**sharafudheen audio (Private)**: 3:03 PM: thank u