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

**NBD-PWD-2016/M0513**

**Source: NBD-PWG**

**Status: Draft**

**Title: Web chat from Meeting of March 22, 2016**

**Chat Log D:\\_wo\1DMG\2015\\_BigDataWG\Docs\ChatLog 2016\_03\_22 14\_59.rtf**

**Lee Anne Davies (to Everyone)**: 1:05 PM: Wo still has a cold!

**Lee Anne Davies (to Everyone)**: 1:06 PM: Where does Frank say to go for colds?

**Tim Zimmerlin (to Everyone)**: 1:18 PM: Mark, the NIST Cloud PWG had a four day workshop. Each day was dedicated to one subject / perspective. It really worked well.

**Tim Zimmerlin (to Everyone)**: 1:20 PM: Mark, instead of a senior editor, I suggest you focus on a formal bibliography as the basis for coherent changes for version 2.

**Wo Chang (to Everyone)**: 1:21 PM: Tim: do they meet every week with four day telecon

**Tim Zimmerlin (to Everyone)**: 1:23 PM: Wo, I attended this workshop one recent Summer. The speakers were well established, including gov, academy, and business. The attendees were global.

**Tim Zimmerlin (to Everyone)**: 1:23 PM: Wo, this workshop was video cast on the Internet.

**Tim Zimmerlin (to Everyone)**: 1:24 PM: Wo, big name speakers will attend only one day.

**Mark Underwood (to Everyone)**: 1:24 PM: Tim et al., worth considering

**Cavan (to Everyone)**: 1:25 PM: I have held back the privacy recommendations, because I want to make sure that a significant number of the commercial sector can use them practically - They could be put out as human procedural goals and aspirations, but I would like it more tied down

**Mark Underwood (to Everyone)**: 1:25 PM: Cavan - yep

**Tim Zimmerlin (to Everyone)**: 1:26 PM: Cavan, I recommend that you consider the financial, regulatory, profit motivations of privacy. You already are an expert, but you must lead the masses who do not care.

**Cavan (to Everyone)**: 1:28 PM: I can work on adding a section that motivates the need and current environment that affects privacy, hopefully that might help people understand what is at stake

**Mark Underwood (to Everyone)**: 1:29 PM: Agree - Wo's suggestion + co-chair curated focus, e.g., container security, or changes to RBAC for big data, traceability across orgs - to help focus third party contribus

**Cavan (to Everyone)**: 1:29 PM: Separately, I think Big Data needs some maturity when it comes to IT security - I am working on that about 1/2 at work

**Tim Zimmerlin (to Everyone)**: 1:31 PM: Cavan, please share documents that you can. I put a lot of emphasis on formal sources.

**Cavan (to Everyone)**: 1:31 PM: For example, understanding the performance trade offs for secure containers, doesn't seem to be readily available.

**Cavan (to Everyone)**: 1:32 PM: Tim - Hopefully, we will have a released write up of the privacy workshop this or next week. It won't cover IT security.

**Tim Zimmerlin (to Everyone)**: 1:32 PM: Cavan, the performance penalty depends on algorithms.

**Cavan (to Everyone)**: 1:33 PM: yes, and also on the kind of jobs being managed I suspect

**Tim Zimmerlin (to Everyone)**: 1:33 PM: Cavan, understand. There are two security tracks. Just the published standards are good enough.

**Tim Zimmerlin (to Everyone)**: 1:35 PM: Cavan, algorithms have been organized by Geoffrey Fox as Ogres. This and UCB Dwarfs are good starting points.

**Ann Racuya-Robbins (to Everyone)**: 1:35 PM: What privacy workshop are you referring to?

**Cavan (to Everyone)**: 1:37 PM: It appears that there is no off the shelf, sending credentials from kerberos -> Atlas -> R ->Spark ->Accumulo -> HDFS; Pieces are done, but from what I understand people are still doing custom glue code to make it all work.

**Cavan (to Everyone)**: 1:38 PM: Anne: MIT & Census put on a Big Data privacy with Industry and Academia last December

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:38 PM: Dear Cavan: Do you have a link web site for follow-up....?Thanks, Pw

**Tim Zimmerlin (to Everyone)**: 1:39 PM: Cavan, I do not know this at all. I usually start with IPsec and ISAkey SW.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:40 PM: Dear Sir: Will folks be able to securily and privately dial into the Face-to-Face via a secure Internet www connection.....?

**Tim Zimmerlin (to Everyone)**: 1:40 PM: As far as I know, Microsoft is the main vendor pushing Kerberos for its distributed services.

**Cavan (to Everyone)**: 1:43 PM: Tim, from what I understand that is insuring that a message is sent to be recieved properly encrypted and decrypted (roughly), I am focusing on insuring that the mechanisms of both authentication and authorization are properly sent through the stack with appropriate logging as would be desired in FISMA

**Cavan (to Everyone)**: 1:43 PM: PW: follow up on what issue?

**Cavan (to Everyone)**: 1:46 PM: Tim, I focus primarily on large scale Linux distributed systems now. So will be testing setting up a separate open-source kerberos server that ties into the LDAP, that synchronizes with the Microsoft X-500 system

**Tim Zimmerlin (to Everyone)**: 1:46 PM: Cavan, understood. Can Certificate Authorities (CAs) and digitally signed documents do the job?

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:47 PM: Dear Cavan: The presentations made at the MIT big data conference.......et al....that's all....also, when conducting a forensics investiagtion tracking down financial fraud....we copy the encrypted data (in real-time, too) then de-crypt our copy to get at the data we need to take the crooks/fraudsters to court for our clients and successifully prosecute.....all the way to clowbacks, restitution and redemption phases of our rules of law.....that's how we do it.....Pw

**Cavan (to Everyone)**: 1:47 PM: Tim, not to my knowledge. We are going to want role based security with detailed views of what a user or role has access to.. more like in a RDB

**Tim Zimmerlin (to Everyone)**: 1:48 PM: Cavan, if you go with Microsoft, Microsoft owns you.

**Tim Zimmerlin (to Everyone)**: 1:48 PM: I have no stake in this. Microsoft is good enough.

**Tim Zimmerlin (to Everyone)**: 1:49 PM: I suggest you research inCommon and GENI, though.

**Cavan (to Everyone)**: 1:49 PM: I agree, just the scale and cost for us is an issue. We are already way to expensive.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:50 PM: Dear Tim: Thanks for the inCommon & GENI suggestions....we'll follow-up with some research....Thanks....

**Tim Zimmerlin (to Everyone)**: 1:50 PM: Cavan, in no manner, do I want to influence your project.

**Tim Zimmerlin (to Everyone)**: 1:50 PM: But, I suggest a pilot environment for some bakeoffs and codefests.

**Cavan (to Everyone)**: 1:50 PM: PW, that is effectively streaming decryption? Interesting set up

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:51 PM: Dear Cavan: Have you already completed a Cost Benefits Analysis.....Risk Assessment....(worst case vs best case)...?

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:52 PM: Dear Cavan: Yes, yes it is......slow and painful......slow and painful.....but clients appreciate the due diligence....Thanks....

**Tim Zimmerlin (to Everyone)**: 1:52 PM: Cavan, have you looked at the government only hyperclouds like AWS? FedRAMP?

**Cavan (to Everyone)**: 1:53 PM: We are trying to get a good technical sense of what really works - we have had almost every vendor and technologist come through to sell us on how to do the 2020 census

**Cavan (to Everyone)**: 1:54 PM: we have a 15 man team evaluating various cloud offerings

**Cavan (to Everyone)**: 1:55 PM: vendors tend to promise the world, we have been burned by internal people not really understanding the technology that they are betting the house on

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:56 PM: Dear Cavan; Set up a dummy 3-day sequential trial session with a follow-up 3-day post trial evaluation....this will sort out the imposters on both side of the table......just a thought....

**Tim Zimmerlin (to Everyone)**: 1:56 PM: Cavan, 2020 Census is just around the corner! I suggest you look into the Obamacare Public Medical Insurance Sites. This is recent history, well documented, Lambda architecture (some prefer Reactive Systems).

**Cavan (to Everyone)**: 1:56 PM: We did and called it a set of DaVinchi sessions

**Cavan (to Everyone)**: 1:57 PM: We had everyone in for about 3 years... in anticipation of the 2020 census. We have to have the system finished by 2018

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:58 PM: Dear Cavan: Two rules.....Vendors lie....and Clients don't know what they need....which makes for some very interesting conversations.....

**Tim Zimmerlin (to Everyone)**: 1:59 PM: Cavan, I rely on digitally signed documents with 2048 bits…as a crucial part of meeting requirments.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:59 PM: Dear Cavan: These are strictly our personal and professional opnions......

**Tim Zimmerlin (to Everyone)**: 2:00 PM: Cavan, in a parallel track, I suggest you standardize on XML documents with XML schemas, including provenance and curation metadata.

**Cavan (to Everyone)**: 2:00 PM: Tim, I agree. I know one of the people who set up Obama care, so I got to hear the blow by blow. Politicians gave them less than a month for testing, used the Verizon cloud (who didn't stress test it)... they didn't use gov't developers... and forced the use of a canadian firm for the authentifican...

**NANCY GRADY (to Everyone)**: 2:00 PM: Sorry all, I need to drop off

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:01 PM: Dear Nancy: Have a nice weekend....

**Cavan (to Everyone)**: 2:03 PM: PW, I agree. And I have been on both sides of the table. And I agree that no customer really knows what they want at the beginning - otherwise there would never be system maintenance and modification.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:09 PM: Dear Sir (aka: Cavan): Honestly, with you being in charge and fully and completely and totally supported and backed-up by those folks in the C-Suites....you can offer up to the project the best way to tackle the vendor issue....beginning with: Budget Resources (both dollars & sense)/Timeline, Requirements Design & Development (a living document that can change over the life of the project), after buy-in the Project Charter is Signed-off and all project documentation is maintained in a secured library for all project members to access and review...) then walk-step the project through phases and gates and gos and no-goes....all the way through to Testing, Implementation & Maintenance and Final Evalutation and Go Live.....We know you can do this with the appropriate support from those folks on-hight.....Just a thought.....

**Tim Zimmerlin (to Everyone)**: 2:11 PM: Ann, please consider the difference between subtractive extracion like coal or oil AND additive extraction like Netflix recommender system.

**Cavan (to Everyone)**: 2:17 PM: PW: the Decennial Directorate has 5 Divisions, one that is dedicated to contracts and procurement. I have no idea of how many certified project managers are running around. ... I wish I were in charge and was the sole voice talking to the C-suites, but I am a simple guy trying to focus on the issues of big data for stat agencies. (But I have directed several large scale development projects, sometimes the project planners have project planners)

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:19 PM: Dear Cavan: You have my utmost respect and empathy.....and our best wishes, too....Respectfully yours, Pw

**Cavan (to Everyone)**: 2:20 PM: PW: your anonymous, boring, gray, bureaucrat thanks you

**Tim Zimmerlin (to Everyone)**: 2:26 PM: Ann, you seem to be distinguishing your physical self from your cyber self: physical space versus cyberspace. I call it your proxy, but other use avatar.

**Tim Zimmerlin (to Everyone)**: 2:27 PM: What leaks in either space also leaks in the other space.

**Tim Zimmerlin (to Everyone)**: 2:28 PM: What about your physical life and cyber life? Does your cyber life continue after death? Do you care?

**Mark Underwood (to Everyone)**: 2:28 PM: The suggestion : to explore how the Quantified Self , augmented reality, wearables to explore where the boundaries of "self" apply for Big Data purposes

**Lee Anne Davies (to Everyone)**: 2:28 PM: I would suggest that for policy or research you should never know who the person is.

**Tim Zimmerlin (to Everyone)**: 2:29 PM: Lee Anne, what about vehicle recalls? What about genomic (or x-omic) risks?

**Lee Anne Davies (to Everyone)**: 2:30 PM: TZ - regarding cyber life after death - there are several lawyers dealing with estates and e-issues. It's a growing problem when settling estates.

**Lee Anne Davies (to Everyone)**: 2:31 PM: TZ - vehicle recall implementation is not policy or research. Why would the researcher need to know that Lee Anne Davies owns car x? They need to know about car x, not me. This is distinct from the mailing out of the recall info.

**Mark Underwood (to Everyone)**: 2:31 PM: Anonymity as degree, not a state

**Tim Zimmerlin (to Everyone)**: 2:31 PM: Lee Anne, are you legally liable for what your avatar does w/o your approval?

**Lee Anne Davies (to Everyone)**: 2:32 PM: I've done a lot of health care research and this is a big issue - I never knew any person for my research. Everything was anonymized.

**Tim Zimmerlin (to Everyone)**: 2:32 PM: Lee Anne, researchers sometimes learn of dangers affecting their subjects. E.g., Flint water supply.

**Lee Anne Davies (to Everyone)**: 2:33 PM: TZ - I don't know the answer to that but we see avatars being misused increasingly in things liked FB, LinkedIn and advertising using people's photos, taken from FB . It's interesting and I think without my explicit or implicit permission I'm not responsible.

**Cavan (to Everyone)**: 2:34 PM: Mark: Anonymity as a degree... is a hot topic. We would love to sell that as an objective. But we haven't been able to articulate what degrees are practical, and haven't been able to sell the public and business on privacy as a degree. Love to hear your ideas

**Lee Anne Davies (to Everyone)**: 2:34 PM: TZ - flint water supply needs to be an independent variable, separate from the individual identifier. Geography, zip codes, etc. but not names, speciifc addresses otherwise it is bias research.

**Tim Zimmerlin (to Everyone)**: 2:34 PM: Lee Anne, AI is good enough (e.g., Eliza of 1960s) to pull off successful masquerades. heh-heh

**Lee Anne Davies (to Everyone)**: 2:35 PM: TZ - What's Eliza? I should google this?

**Lee Anne Davies (to Everyone)**: 2:36 PM: Cavan - I'd enjoy reading it. leeannedavies@agenomics.ca

**Tim Zimmerlin (to Everyone)**: 2:37 PM: Eliza is a very old text based AI system from MIT. Eliza is a pychological therapist. Even when a person knows this, they usually develop a trusting relationship and share private information.

**Lee Anne Davies (to Everyone)**: 2:37 PM: This call just got more interesting!

**Lee Anne Davies (to Everyone)**: 2:38 PM: I'm working at home and my 15 year old just walked by looking oddly at the 'someone looking through your blinds'. How do I explain it is work!

**Tim Zimmerlin (to Everyone)**: 2:38 PM: Yes, social engineering successfully demonstrated by AI.

**Tim Zimmerlin (to Everyone)**: 2:46 PM: Reification is any process of establishing the existence up to identity of an actor, logically inferred from a trail of information.

**Tim Zimmerlin (to Everyone)**: 2:58 PM: Frank, rock - paper -scissors of concerns: privacy, safety, profit.

**Lee Anne Davies (to Everyone)**: 2:58 PM: ok Wo. Looking forward to it.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:59 PM: Dear Sir: Good meeting.....