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

**NBD-PWD-2015/M0487**

**Source: NBD-PWG**

**Status: Draft**

**Title: Web chat from Meeting of Jan. 5, 2016**

**Chat Log D:\\_wo\1DMG\2015\\_BigDataWG\Docs\ChatLog 2016\_01\_05 14\_59.rtf**

**Tim Zimmerlin (to Everyone)**: 1:08 PM: Northeast, South, Midwest, West

**Tim Zimmerlin (to Everyone)**: 1:09 PM: I am talking now.

**Nancy Grady (to Everyone)**: 1:12 PM: I'm on the phone, should be able to join at 1:30

**Geoffrey Fox (to Everyone)**: 1:14 PM: dates ok for me

**Mark Underwood (to Everyone)**: 1:15 PM: Dates OK for me

**Mark Underwood (to Everyone)**: 1:15 PM: 1 vs 2 days will depend on our agenda

**Russell Incognito (to Everyone)**: 1:24 PM: Cavan: please mute

**Mark Underwood (to Everyone)**: 1:37 PM: See http://nyti.ms/1O0f6OK This story reports that Azure is handling a trillion sensor messages weekly.

**Mark Underwood (to Everyone)**: 1:41 PM: A plea to new drafters: please take time to read v1 drafts

**Mark Underwood (to Everyone)**: 1:44 PM: The PII can be provided by private parties, e.g., video surveillance

**Sprague, John D (HQ-JH000) NASA (to Everyone)**: 1:45 PM: whats the goal on having version 2 out? I showed up late

**Mark Underwood (to Everyone)**: 1:49 PM: Agree with Arnab here - it's a matter of degree. If you watch CSI Cyber you'll see a variety of re-identifying info, especially cell data => GPS

**Tim Zimmerlin (to Everyone)**: 1:53 PM: John D, the NBD-PWG has a schedule to release three versions that iteratively improve on the previous work.

**Tim Zimmerlin (to Everyone)**: 1:54 PM: Version 1 was released in steps over last Summer.

**Sprague, John D (HQ-JH000) NASA (to Everyone)**: 1:57 PM: Tim Z, thanks.

**Tim Zimmerlin (to Everyone)**: 2:07 PM: The PII concept is flawed & out of date. Each vertical application has a spectrum of privacy concerns.

**Tim Zimmerlin (to Everyone)**: 2:08 PM: Data aggregators and marketeers have identified most people in the US today. Who is a student? Who is getting married? Who is having a baby? etc.

**Mark Underwood (to Everyone)**: 2:09 PM: +1 Tim, and retailers can gather their own, which work well against public databases with loyalty programs

**Mark Underwood (to Everyone)**: 2:10 PM: There's no silver bullet here, which is why we call for a fabric

**Tim Zimmerlin (to Everyone)**: 2:11 PM: Google and Facebook own data about most US citizens.

**Tim Zimmerlin (to Everyone)**: 2:13 PM: There are two kinds of US datasets, those that have been hacked and those that will be hacked.

**Mark Underwood (to Everyone)**: 2:19 PM: I was on the voter registration DB that was hacked last week

**Tim Zimmerlin (to Everyone)**: 2:19 PM: Highly privileged uses will become the targets and their accounts hijacked. Each security mechanism can be broken over time. The weakest link is the obvious target.

**Mark Underwood (to Everyone)**: 2:19 PM: "not really a hack; just a post of quasi-public data

**Tim Zimmerlin (to Everyone)**: 2:20 PM: Mark, they just exposed the voter information for much of the US. Not really a hack?

**Tim Zimmerlin (to Everyone)**: 2:21 PM: The computer, OS, broswer, cookies are PII.

**Mark Underwood (to Everyone)**: 2:21 PM: Yeah, not technically. There was an aggregator out there who posted a set and later took it down. They would rather have sold it :)

**Tim Zimmerlin (to Everyone)**: 2:22 PM: Mark, what is their punishment?

**Mark Underwood (to Everyone)**: 2:22 PM: None AFAIK, but the FTC might be looking into it

**Tim Zimmerlin (to Everyone)**: 2:23 PM: Mark, what is the difference between a Ninja hacker and a really incompetent system admin?

**Tim Zimmerlin (to Everyone)**: 2:24 PM: The audit logs are provenance data.

**Mark Underwood (to Everyone)**: 2:27 PM: Tim: One of them gets a matching 401(k)

**Tim Zimmerlin (to Everyone)**: 2:28 PM: Marine Corps person, identity provider sanctioning user's actions + well defined user lanes + audit log analysis identifying outliers like day of week, time of day, scheduled vacation, logged in more than once, network latency is too long.

**Tim Zimmerlin (to Everyone)**: 2:29 PM: Mark, hahaha

**Tim Zimmerlin (to Everyone)**: 2:34 PM: Just identifying the distilled concerns of government workers is a reasonable first milestone. Concerns. Relevant experiences.

**Tim Zimmerlin (to Everyone)**: 2:35 PM: Also identifying the relevant laws and policies are worth documenting for CxO executives and directors.

**Mark Underwood (to Everyone)**: 2:38 PM: Cavan - did you loo at Spade yet? Ashish Gehani and Dawood Tariq. 2012. SPADE: support for provenance auditing in distributed environments. In Proceedings of the 13th International Middleware Conference (Middleware '12), Priya Narasimhan and Peter Triantafillou (Eds.). Springer-Verlag New York, Inc., New York, NY, USA, 101-120.

**Cavan Capps (CENSUS) (to Everyone)**: 2:40 PM: Mark - Thanks, I will look it up

**Mark Underwood (to Everyone)**: 2:43 PM: Cavan - it was where V1 left off the with provenance "proposal" - where we'd likely be restarting the conversation

**Cavan Capps (CENSUS) (to Everyone)**: 2:44 PM: Mark - Thanks, love to have stuff already done

**Mark Underwood (to Everyone)**: 2:50 PM: http://bit.ly/1TCFvnf Developing Global Norms for Sharing Data and Results during Public Health Emergencies Pub on PLOS

**Wo Chang (to Everyone)**: 2:54 PM: · <http://www.oracle.com/us/dm/seo100482415-na-us-ip-ipi1-ev-2778233.html>

· https://cs.gmu.edu/~hfoxwell/Symposium/Symp\_bios\_abstracts.html <https://cs.gmu.edu/%7Ehfoxwell/Symposium/Symp\_bios\_abstracts.html>

**Russell Reinsch (to Everyone)**: 2:54 PM: Symposium link

**Mark Underwood (to Everyone)**: 2:58 PM: Hi John- welcome

Hi John- welcome

**Robert Reyling (to Everyone)**: 2:59 PM: Next time will we review the Vol. 2 plans?