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

**NBD-PWD-2015/M0494**

**Source: NBD-PWG**

**Status: Draft**

**Title: Web chat from Meeting of Jan. 26, 2016**

**Chat Log D:\\_wo\1DMG\2015\\_BigDataWG\Docs\ChatLog 2016\_01\_26 15\_03.rtf**

**Tim Zimmerlin (to Everyone)**: 1:09 PM: SnP is confounded by concerns, cross-cutting concerns, mechanisms, and outcomes.

**Tim Zimmerlin (to Everyone)**: 1:11 PM: SnP conversations are dominated by people saying the most recent meme they have heard.

**Russell Reinsch (to Organizer(s) Only)**: 1:13 PM: Wo I am interested in A8 and A14 on the table on your screen , develop mgmt. section and add text from Mark on semantic web and big data.

**Mark Underwood (to Everyone)**: 1:16 PM: This is the Census document on open source of its data - https://github.com/uscensusbureau/open-source-policy

**Mark Underwood (to Everyone)**: 1:24 PM: I think of of Ann's recurring advice is that what Frank is now referring to as "frames" are essentially defined non-computationally. This creates an opening for trust frameworks b/c they have some superficial transparency for lawyers and perhaps the lay public.

**Tim Zimmerlin (to Everyone)**: 1:25 PM: Frank, you are talking about legal concerns here…contracts, rights, recourses.

**Tim Zimmerlin (to Everyone)**: 1:25 PM: Mark, +1

**Ann Racuya-Robbins (to Everyone)**: 1:26 PM: Am I being heard?

**Tim Zimmerlin (to Everyone)**: 1:26 PM: Ann, no.

**Ann Racuya-Robbins (to Everyone)**: 1:26 PM: I would like to speak.

**Ann Racuya-Robbins (to Everyone)**: 1:26 PM: I will relogin

**Tim Zimmerlin (to Everyone)**: 1:29 PM: Is there a difference between public and private data?

**Tim Zimmerlin (to Everyone)**: 1:30 PM: Or is the only difference sources?

**Tim Zimmerlin (to Everyone)**: 1:31 PM: I.e., public data comes from public sources…etc.

**Tim Zimmerlin (to Everyone)**: 1:35 PM: George Mason economics department has generated 2 Nobel prizes. This GM department has a Sim City simulation with 120 million SW agents, one for every US worker today. Soon, they will feed big data into each SW agent and drive US policy.

**Gregory Weidman (to Everyone)**: 1:35 PM: Cavan, you are indicating that there is a piece of the ontology that relates to what a dishonext actor could do with the data.

**Mark Underwood (to Everyone)**: 1:37 PM: Ann - Interesting: "family data" - That does have an ontology mapping. There should be some academic lit on that . . .

**Mark Underwood (to Everyone)**: 1:38 PM: I'm not sure that the increased burden (e.g., reduced response raetes) at Census and elsewhere is an unaccessible outcome

**Mark Underwood (to Everyone)**: 1:38 PM: \*unacceptable

**Gregory Weidman (to Everyone)**: 1:38 PM: Cavan, I think you are going the wrong direction on the privacy problem. More interesting is situations where you can filter by behavior, and then resolve identified behavior and associate that with an individual person.

**Tim Zimmerlin (to Everyone)**: 1:40 PM: Frank's frame based reasoning was published in the 1980's by MIT's Minsky and his students. It was one of three AI approaches, including expert systems and logic programming. All three live on today.

**Tim Zimmerlin (to Everyone)**: 1:41 PM: In mem. Marvin Minsky...

**Gregory Weidman (to Everyone)**: 1:41 PM: RIP

**Gregory Weidman (to Everyone)**: 1:44 PM: An important question of the "Frame" is the question of "who is already inside the box"? This gets to Ann's point about "family data." Even if we can identify the frame, and can identify the expectations that the user has about the frame, we are then challenged to identify who else is inside that frame, and how much rights they have to break the frame.

**Tim Zimmerlin (to Everyone)**: 1:47 PM: Ann, there is revenue sharing today. Members get choice to share member id & get discounts. Psychologists understand people trade on privacy for personal benefits.

**Mark Underwood (to Everyone)**: 1:47 PM: Did anyone else watch the FTC Privacy Conference on Jan 23?

**Ann Racuya-Robbins (to Everyone)**: 1:47 PM: yes

**Cavan (to Everyone)**: 1:48 PM: I read summary of it

**Cavan (to Everyone)**: 1:48 PM: Greg: have to think about how to reply to your point

**Tim Zimmerlin (to Everyone)**: 1:49 PM: An obvious outcome of big data in the 21st Century is the creation of open marketplaces and closed clearinghouses for data globally.

**Mark Underwood (to Everyone)**: 1:50 PM: Sorry Jan 21 - was a good overview of academic approaches to these issues, incl. software agents and risks for unfair or faulty analytics / deep learning outcomes

**Tim Zimmerlin (to Everyone)**: 1:50 PM: GM University runs a SW agent for each one of us.

**Tim Zimmerlin (to Everyone)**: 1:51 PM: I wonder if my SW agent is doing better than me.

**Cavan (to Everyone)**: 1:52 PM: I know my SW agent is doing better..

**Gregory Weidman (to Everyone)**: 1:53 PM: LOL!

**Tim Zimmerlin (to Everyone)**: 1:53 PM: I want to get advice from mine.

**Russell Reinsch (to Everyone)**: 1:53 PM: Tim - probably; just like the GDP and CPI are better than their real \_

**Mark Underwood (to Everyone)**: 1:53 PM: The FTC workshop was recorded at http://1.usa.gov/1K8Lkc8 See #privacycon for some complaints from the field about what was covered

**Cavan (to Everyone)**: 1:54 PM: thx for the link

**Ann Racuya-Robbins (to Everyone)**: 1:56 PM: Fair enough Cavan for census

**Mark Underwood (to Everyone)**: 1:56 PM: I'm dubious that any of these use cases are substantively different from the ones we already presented in V1

**Mark Underwood (to Everyone)**: 1:57 PM: Moving ahead means deeper mapping to RA - e.g., how apps get built, and that's where we can try to get better mindshare among implementers

**Wo Chang (to Everyone)**: 1:58 PM: Mark: can you ref/cite which Vol. 4 SnP sections that are related to this conversation?

**Tim Zimmerlin (to Everyone)**: 2:10 PM: Mark, I bet statistical agencies possess sufficient provenance data. I am considering best practices here.

**Cavan (to Everyone)**: 2:10 PM: Mark speaks my mind

**Ann Racuya-Robbins (to Everyone)**: 2:11 PM: Thanks Mark

**Cavan (to Everyone)**: 2:11 PM: @Tim: we could and should always improve

**Tim Zimmerlin (to Everyone)**: 2:13 PM: Cavan, track the cell phones of the surveyors doing field polling, using an installed app.

**Cavan (to Everyone)**: 2:13 PM: that we do

**Cavan (to Everyone)**: 2:14 PM: with advice from UPS

**Tim Zimmerlin (to Everyone)**: 2:19 PM: Privacy has no technology solution, just legal and contractual ones. Tech could help or hurt things.

**Ann Racuya-Robbins (to Everyone)**: 2:20 PM: My question has to do with "operational requirements of the system" that could be anything...is that really a standardizing concept?

**Tim Zimmerlin (to Everyone)**: 2:21 PM: Ann, yes, it is important. E.g., record keeping, audits, breach notification, opt-out, etc.

**Ann Racuya-Robbins (to Everyone)**: 2:22 PM: Why are we still tying to PII which is not well understood.

**Tim Zimmerlin (to Everyone)**: 2:23 PM: safe harbors, due diligence claims, ...

**Ann Racuya-Robbins (to Everyone)**: 2:23 PM: Does this privacy fabric contain EU GDPR?

**Mark Underwood (to Everyone)**: 2:27 PM: Missing in the Underwood doc - requirements that can be mapped to conformity

**Gregory Weidman (to Everyone)**: 2:29 PM: Ringing off at 2:30. Good meeting, everyone!

**Tim Zimmerlin (to Everyone)**: 2:35 PM: ACLs do not translate to external systems. Federation is required to communication ACLs formally.

**Tim Zimmerlin (to Everyone)**: 2:36 PM: Identity providers, indentity managers, trust managers are required for ACLs to work. Also, man in the middle attack vectors are everywhere here.

**Tim Zimmerlin (to Everyone)**: 2:39 PM: Opaque identifiers are a failure mode.

**Ann Racuya-Robbins (to Everyone)**: 2:39 PM: Good point Mark!

**Tim Zimmerlin (to Everyone)**: 2:40 PM: IPsec with digital signatures are necessary. This brings in IdP, IdM, trust, CA/RA, white/black lists. This is way bigger than this conversation.

**Mark Underwood (to Everyone)**: 2:43 PM: Frank: some value is coming from the aviation reporting system. Is there an equivalent for S&P? (The fTC workshop did not arrive at a single answer, but touched on it(.

**Mark Underwood (to Everyone)**: 2:47 PM: Tim-, yes = are there federation standards that map to our work?

**Mark Underwood (to Everyone)**: 2:47 PM: Beyond NIST CSA

**Tim Zimmerlin (to Everyone)**: 2:48 PM: Mark, there are working examples today. IPsec as supported by Java libraries (and many others) are used.

**Tim Zimmerlin (to Everyone)**: 2:48 PM: Clouds, microservice, OpenStack, SOAP, REST, etc.

**Tim Zimmerlin (to Everyone)**: 2:49 PM: Internet2, DOE & NSF supercomputing centers, etc.

**Mark Underwood (to Everyone)**: 2:51 PM: Tim, the orchestration notions I've seen from some of those have little to no data-dependent role specifications; the security models don't map to data, they map to some feature of the Apache program that is being implemented, which is too narrow for both SnP

**Mark Underwood (to Everyone)**: 2:52 PM: There's that data-driven model in academia that perhaps we should revisit

**Tim Zimmerlin (to Everyone)**: 2:53 PM: Mark, yes and no. Authentication, authorization, and auditing can be federated with certs & creds adjudicated by IdPs, IdMs, and trust providers.

**Cavan (to Everyone)**: 2:54 PM: have to go, next meeting

**Mark Underwood (to Everyone)**: 2:55 PM: Tim, understood. But if the developers don't understand the issues when the app is built, these methods can fail to protect beyond a global level. The variety challenge is often missed in big data SDLC

**Tim Zimmerlin (to Everyone)**: 2:56 PM: Data access solutions are application dependent, especially vertical market dependent.

**Russell Reinsch (to Everyone)**: 2:56 PM: I think developing more on the Management definition for v2 is a worthy activity

**Tim Zimmerlin (to Everyone)**: 2:57 PM: Civilian users could be forced to use curated views with tools for data & pattern discovery and visualization.

**Mark Underwood (to Everyone)**: 2:58 PM: Insider threat (HIPAA privacy violations are a special example of this) often exposes failures in SnP SDLC

**Tim Zimmerlin (to Everyone)**: 2:59 PM: Mark, auditing should catch insider threats. NOCs are necessary.

**Tim Zimmerlin (to Everyone)**: 2:59 PM: And least privilege, hours of work, login location, etc.

**Ann Racuya-Robbins (to Everyone)**: 2:59 PM: Have to leave now. Thanks all.

**Mark Underwood (to Everyone)**: 3:01 PM: Tim, that's a good one for a future meeting - "Is auditing enough"

**Tim Zimmerlin (to Everyone)**: 3:02 PM: With IdPs, IdMs, trust managers, yes. Rely on creds & creds. Block ssh for all uses except emergencies.

**Tim Zimmerlin (to Everyone)**: 3:02 PM: error…creds & certs

**Tim Zimmerlin (to Everyone)**: 3:02 PM: Include live NOC approvals at critical points.

**Mark Underwood (to Everyone)**: 3:03 PM: Thanks for being here, everyone!