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

**NBD-PWD-2015/M0484**

**Source: NBD-PWG SnP Subgroup**

**Status: Draft**

**Title: Web chat from SnP Subgroup Meeting of Dec. 16, 2015**

**Chat Log C:\underwood\collaboration\GoToMeeting\ChatLog NBD\_PWG 2015\_12\_16 17\_03.rtf**

**Russell Reinsch (to Everyone)**: 4:14 PM: Notes: reidentify. secondary attributes. data at aggregators not tracked. no standard snp techniques. proposal --> way to audit.

**Mark Underwood (to Everyone)**: 4:16 PM: Recall "conformance" conversation w/ Frnak

**Mark Underwood (to Everyone)**: 4:16 PM: \*Frank

**Russell Reinsch (to Everyone)**: 4:17 PM: three parts. 1. take PII in separate store. 2. link. what you bought separate from who you are. agreed to provenence doc for tracking. 3. log.

**Russell Reinsch (to Everyone)**: 4:17 PM: log access, activity on DB

**Mark Underwood (to Everyone)**: 4:18 PM: Cavan is also concerned w/ access via analytics

**Russell Reinsch (to Everyone)**: 4:20 PM: how to distiquish manual extraction, which may look benign. not currently solution for.

**Mark Underwood (to Everyone)**: 4:20 PM: The audit issue is not technologically different from audit from systems management

**Tim Zimmerlin (to Everyone)**: 4:20 PM: There are few semi-formal methods to enforce SnP. One way starts by defining "what are the risks", "who are the threats", "what are the counter measures".

**Mark Underwood (to Everyone)**: 4:21 PM: Arnab: A formal threat model s/b articulated

**Tim Zimmerlin (to Everyone)**: 4:23 PM: The financial community does its SnP; however, their "big enforcement methodology" is to undo the damage and punish the perpetrators.

**Tim Zimmerlin (to Everyone)**: 4:24 PM: My favored last ditch SnP methodology is real time monitoring and rapid reaction to stop the bad actors.

**Russell Reinsch (to Everyone)**: 4:24 PM: EU not want privacy included (ISO side). More of a NIST line

**Mark Underwood (to Everyone)**: 4:26 PM: Tim provided a couple of citations about privacy work in EU

**Tim Zimmerlin (to Everyone)**: 4:26 PM: My favored first line SnP methodology is training backed by policing.

**Mark Underwood (to Everyone)**: 4:28 PM: Cavan: Privacy concerns in the US are reducing the response rates somewhat dramatically

**Tim Zimmerlin (to Everyone)**: 4:29 PM: The EU has recently passed a new privacy law with teeth: http://www.bbc.com/news/technology-35110909

**Mark Underwood (to Everyone)**: 4:30 PM: Cavan: Big Data is relevant here b/c new data sources may be needed in order to make new inferences or to make them sooner

**Russell Reinsch (to Everyone)**: 4:30 PM: challenge: to build confidential dataset. graph / network analysis - need PII.

**Mark Underwood (to Everyone)**: 4:34 PM: V1 discusses provenance tracking in some detail

**Tim Zimmerlin (to Everyone)**: 4:34 PM: Cavan: Census could use transparency, including how the big data helps citizens.

**Russell Reinsch (to Everyone)**: 4:35 PM: Collect: how businesses use data.

**Mark Underwood (to Everyone)**: 4:35 PM: Cavan: Would be good to solicit private firm response to whatever might be proposed

**Tim Zimmerlin (to Everyone)**: 4:36 PM: Cavan: Census could use "opt in".

**Mark Underwood (to Everyone)**: 4:37 PM: Mark U: Marketing is kinda moving in an opposite direction; Pepper and Rogers gave it a name: the One toOne future

**Mark Underwood (to Everyone)**: 4:37 PM: which means non-statistical marketing

**Mark Underwood (to Everyone)**: 4:40 PM: In the US, the penalties are in the Fair Credit Reporting Act and the HIPAA violations

**Mark Underwood (to Everyone)**: 4:43 PM: There are HIPAA "barns" - Some web hosts tell u they are explicitly not HIPAA compliant - and this is privacy related.

**Mark Underwood (to Everyone)**: 4:45 PM: Tim: online policing and filtering

**Arnab Roy (to Everyone)**: 4:45 PM: barns could be constructed using more sophisticated encryption systems, like, homomorphic and functional, which allow computation on encrypted data. Currently these are very inefficient but I am hopeful they could be mainstream within 10-15 years.

**Mark Underwood (to Everyone)**: 4:46 PM: +1 Tim - it's a "framework" solution which takes into account the types of tools available to attack, detect, remediate

**Mark Underwood (to Everyone)**: 4:47 PM: We have had a healthy debate about what to do wtih provenance in general as part of the V1 document struggle; it's not only a privacy challenge

**Mark Underwood (to Everyone)**: 4:56 PM: - We will need to make some headway on the PII definition - operationalization problem

**Russell Reinsch (to Everyone)**: 4:56 PM: Challenge: define PII.

**Mark Underwood (to Everyone)**: 4:57 PM: That's the FTC RAdio Shack ruling - the new buyers wanted to use the data differently than was agreed to

**Mark Underwood (to Everyone)**: 5:00 PM: Cavan: warns that overly complicated solutions such as DDI could fall short

**Mark Underwood (to Everyone)**: 5:01 PM: Arnab: more nuanced goals to judge where utility of data is overly constrained

**Russell Reinsch (to Everyone)**: 5:01 PM: Expand goals. Does a proposed system meet goals (Arnab).

**Russell Reinsch (to Everyone)**: 5:01 PM: Enforcement of whatever people agreed to; and provide auditing.

**Russell Reinsch (to Everyone)**: 5:02 PM: Some people have already agreed to comment.

**Mark Underwood (to Everyone)**: 5:03 PM: Can is hoping to have something more by week2 of Jan if possible

**Mark Underwood (to Everyone)**: 5:03 PM: \*Cavan