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

**NBD-PWD-2016/M0557**

**Source: NBD-PWG**

**Status: Draft**

**Title: Web chat from Meeting of August 30, 2016**

**Chat Log D:\\_wo\1DMG\2015\\_BigDataWG\Docs\ChatLog 2016\_08\_30 15\_10.rtf**

**Mark Underwood (to Everyone)**: 1:06 PM: Wo - At some point need to discuss feedback from NASA / Cotnoir, Marc

**Mark Underwood (to Everyone)**: 1:17 PM: DODInst

**Cavan Paul Capps (to Everyone)**: 1:22 PM: Sounds like Claude Shannon's, reliability of information exchange

**Cavan Paul Capps (to Everyone)**: 1:25 PM: may I ask a clarifying question?

**Mark Underwood (to Everyone)**: 1:29 PM: Mike - See also ISE.gov - We follow this as a potential family of use cases The site isn't responding at the moment or I would send a link

**Mark Underwood (to Everyone)**: 1:30 PM: Snip from my last email from them "The SBU Technical Advisory Committee (STAC) has been chartered to promote the sharing of terrorism-related information within and between SBU environments. Mr. Alan Rosenhauer, Chief Technology Officer for Regional Information Sharing Systems (RISS) serves as the chair of the STAC.

The STAC’s mission is to advance the timely and secure sharing of SBU criminal justice and terrorism-related information across a wide spectrum of stakeholder missions and security domains, with a focus on Identity, Credential, and Access Management (ICAM). The initial members of the STAC include the Federal Bureau of Investigation (FBI) Criminal Justice Information Services Division, the intelligence community’s Intelink, the U.S. Department of Homeland Security’s (DHS) Homeland Security Information Network (HSIN), and RISS."

**Mark Underwood (to Everyone)**: 1:30 PM: It's not "big data" per se, but we are examining it for federation, partial data sharing and scalability

**Mark Underwood (to Everyone)**: 1:32 PM: Nancy's comment ("What's 'Big Data' about it?") is a technical prophylactic that we use in the group to help limit our scope

**Mark Underwood (to Everyone)**: 1:32 PM: Mike - What is the DodInst # again?

**Staufenberg, Michael [USA] (to Everyone)**: 1:35 PM: DOD Instruction 8330.01 dated May 21, 2014

**Mark Underwood (to Everyone)**: 1:36 PM: thx

**Mark Underwood (to Everyone)**: 1:47 PM: Good to recall that the VA is DoD

**Staufenberg, Michael [USA] (to Everyone)**: 1:49 PM: VA is anightmare

**Staufenberg, Michael [USA] (to Everyone)**: 1:49 PM: from a records management perspective

**Mark Underwood (to Everyone)**: 1:51 PM: Yep, but also on the forefront of R&D - e.g., Mitre collaboration; a brutal irony

**Staufenberg, Michael [USA] (to Everyone)**: 1:52 PM: i work with mitre take the fifth

**Mark Underwood (to Everyone)**: 1:52 PM: lol

**Mark Underwood (to Everyone)**: 1:56 PM: This sounds proscriptive, not descriptive

**Cavan Paul Capps (to Everyone)**: 2:07 PM: I would like to make a comment on this definition of privacy

**Mark Underwood (to Everyone)**: 2:14 PM: In the SnP subgroup we are wrestling w/ how / if to integrate Communicator and Human Subject into existing UML models; the problem is that existing frameworks think they have already handled this via roles & actors, and "communicator" as screens and reports. To be specific in a big data setting, the v2 Vol4 faces some hurdles

**Staufenberg, Michael [USA] (to Everyone)**: 2:25 PM: anyone familar with NIEM?

**Staufenberg, Michael [USA] (to Everyone)**: 2:26 PM: anyone familar with niem?

**Tim Zimmerlin (to Everyone)**: 2:27 PM: Yep, several people know something.

**Tim Zimmerlin (to Everyone)**: 2:27 PM: ...mainly federated sharing aspects.

**Staufenberg, Michael [USA] (to Everyone)**: 2:28 PM: yes - we are exploring common profile right now

**Tim Zimmerlin (to Everyone)**: 2:28 PM: Easy to start with SARs and take small safe steps.

**Mark Underwood (to Everyone)**: 2:31 PM: Yep, familiar with NIEM - long conversation there

**Staufenberg, Michael [USA] (to Everyone)**: 2:32 PM: lots of debate here about niem

**NANCY GRADY (to Everyone)**: 2:32 PM: I'd be interested in the conversation, since a coworker spent a long time working on NIEM

**Mark Underwood (to Everyone)**: 2:34 PM: I wrote the wikipedia page on UCORE

**Tim Zimmerlin (to Everyone)**: 2:36 PM: The newest trends are disruptions. Linux unikernel is a tiny OS. GPUs; FPGAs: ASICs necessary for deep learning. Deep memory hierarchy is not compatible with simple vectorized instructions. C/C++/Java is far removed from hardware.

**Tim Zimmerlin (to Everyone)**: 2:37 PM: The Hadoop Distributed File System (HDFS) is the only data centric constant in the Ref Architecture.

**Cavan Paul Capps (to Everyone)**: 2:37 PM: Tim, know deep learning, what is deep memory

**Tim Zimmerlin (to Everyone)**: 2:39 PM: Cavan, today's newest chips have three levels of SDRAM caches, many cores each with many registers, extensive bus sniffing for stale value detention (like dirty bits used for disk sectors).

**Cavan Paul Capps (to Everyone)**: 2:40 PM: yes

**Cavan Paul Capps (to Everyone)**: 2:40 PM: cores use the same fast and slow cache, is that right?

**Tim Zimmerlin (to Everyone)**: 2:41 PM: ...and mass storage now mixes Flash with rotating HDDs 1000 times slower...depending on where data value resides, your instruction time is very unpredictable.

**Cavan Paul Capps (to Everyone)**: 2:41 PM: ahh.. thx

**Tim Zimmerlin (to Everyone)**: 2:42 PM: Cavan, the worst performance harm is all the superscalar instruction stalls. Very unpredictable. Difficult decision on whether to retry the instruction or simply wait.

**Cavan Paul Capps (to Everyone)**: 2:43 PM: interesting, thanks

**Tim Zimmerlin (to Everyone)**: 3:00 PM: Mark, I agree. Each domain needs to be designed separately.

**Tim Zimmerlin (to Everyone)**: 3:01 PM: Each domain needs its own vocabulary and semantics, not incumbered by others.

**Tim Zimmerlin (to Everyone)**: 3:02 PM: Ok...

**Tim Zimmerlin (to Everyone)**: 3:04 PM: UML, Eclipse, DSLs

**Tim Zimmerlin (to Everyone)**: 3:04 PM: SQL, noSQL, newSQL, Lambda arch

**Wo Chang (to Mark Underwood)**: 3:04 PM: can we talk over the phone after today's call?

**Tim Zimmerlin (to Everyone)**: 3:05 PM: Open Stack, AWS, GCE, Azure, IBM Watson, Tensorflow,...

**Tim Zimmerlin (to Everyone)**: 3:05 PM: HDSF, Spark, Puppet, Yarn, Zookeeper

**Tim Zimmerlin (to Everyone)**: 3:06 PM: ...whoops...HDFS

**Tim Zimmerlin (to Everyone)**: 3:08 PM: Signing off now...