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

**NBD-PWD-2017/M0612**

**Source: NBD-PWG**

**Status: Draft**

**Title: Web chat from Meeting of March 28, 2017**

**Chat Log D:\\_wo\1DMG\2015\\_BigDataWG\Docs\ChatLog 2017\_03\_28 15\_08.rtf**

**Ann Racuya-Robbins (to Everyone)**: 1:22 PM: the pdf option is reasonable

**Geoffrey Fox (to Everyone)**: 1:34 PM: http://www.exascale.org/bdec/sites/www.exascale.org.bdec/files/Baidu\_ABC\_Platform.pdf baidu cloud and hpc, big data

**Tim Zimmerlin (to Everyone)**: 1:46 PM: In social media, not valid data is common. Amazon.com has fake customer evaluations, Twitter has many chatbots, Facebook has people vandalizing others' facebook pages, Wikipedia has dualing chatbots fighting each other (machine vs machine).

**Tim Zimmerlin (to Everyone)**: 1:46 PM: ...on the Interent, the machine wins.

**Nancy Grady (to Everyone)**: 1:47 PM: @Tim I think you've just described "fake data"

**Laurie Aldape (to Everyone)**: 1:47 PM: Please collect references to HPC, cloud, IoT, and CPS to include in Volume 3. Send the references to Nancy and Laurie.

**Tim Zimmerlin (to Everyone)**: 1:47 PM: ...EHRs are trivially easy to modify with complete lies...a lifelong non smoker becomes a heavy smoker. etc.

**Tim Zimmerlin (to Everyone)**: 1:48 PM: Nancy, "fake data" is fake data...all the way down.

**Tim Zimmerlin (to Everyone)**: 1:50 PM: Frank, I have a seasonal allergy. The Internet pollen sites are often completely non valid, contradicting each other, just wrong.

**Tim Zimmerlin (to Everyone)**: 1:50 PM: ...the pollen levels, the pollen sources, today are completely different.

**Tim Zimmerlin (to Everyone)**: 1:54 PM: US experts relate their experiences with the people performing polls, the impact of daily news cycles, and more on even the most carefully designed surveys.

**Tim Zimmerlin (to Everyone)**: 1:55 PM: ...big data is messy, biased.

**Tim Zimmerlin (to Everyone)**: 1:56 PM: ...the better polls ask the same question multiple times, from diverse perspectives, surrounded by neutral questions.

**Ann Racuya-Robbins (to Organizer(s) Only)**: 1:56 PM: Can you make me the presenter?

**Tim Zimmerlin (to Everyone)**: 1:59 PM: I suggest that ALL V WORDS are pain points, negatives, the opposite of positives, risks, failure modes.

**Tim Zimmerlin (to Everyone)**: 2:02 PM: ...V WORDS frame a non technical narrative of failures.

**Cavan (to Everyone)**: 2:11 PM: is this working? Having technical issues

**Tim Zimmerlin (to Everyone)**: 2:12 PM: Cavan, if you are speaking, we cannot hear you now.

**Cavan (to Everyone)**: 2:12 PM: seems that my mike isn't working, thanks Tim

**Cavan (to Everyone)**: 2:14 PM: I still have an issue with emergence

**Cavan (to Everyone)**: 2:15 PM: that seems to speak about self organization in complex adaptive systems

**Cavan (to Everyone)**: 2:16 PM: it isn't important, but the stat agencies typically use the term the mosaic effect to describe the re-identification of data by using multiple data sets together.... just a thought

**Bob Marcus (to Everyone)**: 2:19 PM: hubris = vanity :>)

**Tim Zimmerlin (to Everyone)**: 2:19 PM: Cavan, several good points. E.g., emergence in Wikipedia, Facebook, Twitter, and other Internet social media.

**Cavan (to Everyone)**: 2:19 PM: yes

**Tim Zimmerlin (to Everyone)**: 2:20 PM: Bob, Greek hubris equate "man to god" and evokes the wrath of the gods.

**Tim Zimmerlin (to Everyone)**: 2:20 PM: ...today, management professors teach luck vs skill in business outcomes.

**Cavan (to Everyone)**: 2:22 PM: Frank, don't tell Congress

**Tim Zimmerlin (to Everyone)**: 2:22 PM: ...be wary of "data worshippers".

**Cavan (to Everyone)**: 2:22 PM: We aren't allowed to sample for decennial

**Cavan (to Everyone)**: 2:23 PM: thanks

**Tim Zimmerlin (to Everyone)**: 2:25 PM: Cavan, social graphs represent "emergence".

**Tim Zimmerlin (to Everyone)**: 2:25 PM: ...ANOVA does not.

**Frank Farance (to Everyone)**: 2:25 PM: Cavan, just trying to get that point out, but to explain in it a way that does NOT touch upon political parties, but framing the differences between Quality Management in Physical Objects (e.g., productioin model applies) vs. Information Objects (e.g., productiion model might not apply, sampling technique might produce higher quality data).

**Frank Farance (to Everyone)**: 2:31 PM: FYI, the problem with the 2010 census in New York City (which I was able to isolate here on Roosevelt Island because (1) we're all high rises, and (2) we actually have access to tenants lists in buildings) is that the Census Bureau survey methodology has problems when surveying high-rise buildings, which apply (typically) only in very large cities. For example, there were strong undercounts in Manhattan (lots of high rises) and Roosevelt Island (all high rises). The way we proved this was that we were able to actually count the people on Roosevelt Island (due to quirks in the NY State administration of this City property). It's not about Democrats vs. Republicans, it's about surevey detached houses and garden apartments (low rises) versus high-density high rises (10+ floors).

**Cavan (to Everyone)**: 2:31 PM: Tim, yes social graphs do document a type of emergence. (but the classic examples of emergence is often unofficial adoption of best practice as unofficial standard on the internet, and of course repeated solutions for biological organization in evolution... as I understand it. I hadn't really thought of social graphs as a documentation of emergence... interesting

**Tim Zimmerlin (to Everyone)**: 2:32 PM: Cavan, the antithesis of emergence is stability...2 POVs of a system.

**Tim Zimmerlin (to Everyone)**: 2:33 PM: ...metrics & measurement processes that simultaneously address both sides...I tend to rely on stability.

**Cavan (to Everyone)**: 2:34 PM: Frank, I think the Census agreed with you for the 2000? census. I think that Congress disagreed. So currently the legal definition of a complete enumeration is a count... very expensive.

**Tim Zimmerlin (to Everyone)**: 2:34 PM: ...e.g., 2006 Flash Crash followed by 2007 Liquidity Crisis.

**Bob Marcus (to Everyone)**: 2:41 PM: Smart Cities are an important Use Case for Big Data. See recent NIST documents at https://pages.nist.gov/smartcitiesarchitecture/

**Russ Reinsch (to Everyone)**: 2:43 PM: Go back and consider column identifiers for table 4

**Tim Zimmerlin (to Everyone)**: 2:43 PM: Bob, are you prepared to share your knowledge of smart X, IoT, CPS?

**Tim Zimmerlin (to Everyone)**: 2:44 PM: ...I am researching 5G as the central, common tech.

**Tim Zimmerlin (to Everyone)**: 2:45 PM: ...today's hyperclouds are threatened by FOG computing and Edge Computing".

**Tim Zimmerlin (to Everyone)**: 2:47 PM: ...around 19 critical US infrastructures have been identified.

**Bob Marcus (to Everyone)**: 2:47 PM: Tim: I will be working on Big Data for Smart Cities during the next few months. I think it will be a leading edge application area combining IoT, Big Data, and Cloud + There is a major NIST and international initiative on Smart Cities

**Robert Reyling (to Everyone)**: 2:49 PM: Yes, please post Russ's latest version!

**Tim Zimmerlin (to Everyone)**: 2:52 PM: Bob, I talked to the Barcelona smart city group in DC. I have listened to AT&T and Verizon on 5G. I am reading National Instruments tech reports on 5G and smart X.

**Tim Zimmerlin (to Everyone)**: 2:53 PM: ...TDD, FDD, SDD, SCDD for incredibly dense RANs.

**Tim Zimmerlin (to Everyone)**: 2:53 PM: ...Massive MIMO.

**Tim Zimmerlin (to Everyone)**: 2:54 PM: ...street canyons.

**Tim Zimmerlin (to Everyone)**: 2:55 PM: ...personally, I do like smart parking incl reserved parking.

**Tim Zimmerlin (to Everyone)**: 2:57 PM: ...channel aggregation and mgnt.

**Tim Zimmerlin (to Everyone)**: 2:57 PM: ...the first loser is real money, all about credit cards. Bankers win.

**Ann Racuya-Robbins (to Everyone)**: 2:58 PM: 2:54 PM

Tim Zimmerlin <<A HREF=>

Standards by perspective allows for gaps and emergence to be addressed

**Tim Zimmerlin (to Everyone)**: 2:58 PM: ...that ensures successful outcomes.

**Tim Zimmerlin (to Everyone)**: 2:59 PM: Ann, I do not understand...

**Ann Racuya-Robbins (to Everyone)**: 3:00 PM: sorry Tim this was meant for everyone and Russell

**Tim Zimmerlin (to Everyone)**: 3:00 PM: ...standards are great...everyone has their own. At university, this comment was raised to tone down standards debates.

**Cavan (to Everyone)**: 3:01 PM: Tim, I think of emergence coming from complex adaptive systems as developing what I would call "fuzzy standards"

**Cavan (to Everyone)**: 3:03 PM: have another meeting.. bye and thanks

**Russ Reinsch (to Everyone)**: 3:04 PM: examples at the leaves of the trees vol 9. figure 1. [on premise, cloud, ...]

**Ann Racuya-Robbins (to Everyone)**: 3:04 PM: Good work Russell

**Tim Zimmerlin (to Everyone)**: 3:05 PM: Cavan, for the NBD-PWG, I consider emergence in marketplaces as central.

**Tim Zimmerlin (to Everyone)**: 3:06 PM: ...marketplaces supercharge commerce, digitization, data flows.

**Tim Zimmerlin (to Everyone)**: 3:06 PM: ...social networks, e commerce, MOOCs, news...