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

**NBD-PWD-2016/M0547**

**Source: NBD-PWG**

**Status: Draft**

**Title: Web chat from Meeting of July 26, 2016**

**Chat Log D:\\_wo\1DMG\2015\\_BigDataWG\Docs\ChatLog 2016\_07\_26 15\_09.rtf**

**David W. Boyd (to Everyone)**: 1:28 PM: Can I get in the queue

**Ann Racuya-Robbins (to Everyone)**: 1:31 PM: Thank Frank

**Mark Underwood (to Everyone)**: 1:31 PM: For more reading on this: https://groups.google.com/forum/#!searchin/ontolog-forum/taxonomy/ontolog-forum/RsYMOsTLeRg/e7\_2SATGEQAJ

**Mark Underwood (to Everyone)**: 1:33 PM: Some may find the SKOS practice at W3C helpful: https://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/

**Tim Zimmerlin (to Everyone)**: 1:35 PM: Taxonomy is to Science AS Ontology is to Scientology.

**Tim Zimmerlin (to Everyone)**: 1:36 PM: ...think about that...take your time...have a beer...

**Cavan (to Everyone)**: 1:37 PM: @Tim: lol

**Mark Underwood (to Everyone)**: 1:39 PM: lol Tim - Shades of Kimmy Shimmel http://www.bustle.com/articles/155364-is-cosmetology-real-on-unbreakable-kimmy-schmidt-it-may-seem-familiar

**Mark Underwood (to Everyone)**: 1:40 PM: \*Schmidt

**Tim Zimmerlin (to Everyone)**: 1:41 PM: Mark, read your link...yep...Scientology is a real leader in ontology.

**Mark Underwood (to Everyone)**: 1:43 PM: Frank - Thx for that snip from s is from ISO/IEC Guide 2. Interesting separation of "process"

**Mark Underwood (to Everyone)**: 1:43 PM: \*snip from . . ."

**Tim Zimmerlin (to Everyone)**: 1:44 PM: Russell, I recommend Dwarfs, Giants, Ogres instead of Booze's taxonomy. I read the Booze report previously...it is good but UC Berkeley & Prof. Geoffrey Fox are years ahead.

**Tim Zimmerlin (to Everyone)**: 1:45 PM: Nancy, Russell is looking at BD analytics.

**Mark Underwood (to Everyone)**: 1:46 PM: +1 Nancy. Scoping challenge is [sic] huge

**Mark Underwood (to Everyone)**: 1:49 PM: I'd suggest thinking of these representations as graphs vs. hierarchies, as noted in the SKOS doc

**Tim Zimmerlin (to Everyone)**: 1:49 PM: National Academies starts with hardware, software, and algorithms. I.e., distributed "x"

**Tim Zimmerlin (to Everyone)**: 1:50 PM: Russell, mixes all three indiscriminately at times.

**Tim Zimmerlin (to Everyone)**: 1:51 PM: David Patterson starts with compute, storage, and networks. I.e., Patterson is a big hardware architecture pro...RAID,,,,RISC,,,,Open Compute Risc V.

**Tim Zimmerlin (to Everyone)**: 1:52 PM: Geoffrey Fox starts with his MOOC...around 30 facets of big data systems and analytics.

**Tim Zimmerlin (to Everyone)**: 1:53 PM: ....Geoffrey relies on use cases to verify and validate his taxons.

**Mark Underwood (to Everyone)**: 1:53 PM: One helpful existing class scheme, which I mentioned long ago is the ACM one used for ACM content classification

**Mark Underwood (to Everyone)**: 1:53 PM: +1 for use case driven approach; operationalize vs. defie

**Mark Underwood (to Everyone)**: 1:54 PM: \*define

**Tim Zimmerlin (to Everyone)**: 1:54 PM: Mark, yes!

**Tim Zimmerlin (to Everyone)**: 1:56 PM: Big data is an opaque "buzz word". This makes big data a lightening rods for disagreement.

**Tim Zimmerlin (to Everyone)**: 1:57 PM: Big data begs for a third word to complete the context...speeds & feeds...architectures...analytics...custom hardware, etc.

**NANCY GRADY (to Everyone)**: 2:00 PM: sorry, I need to step out...

**Mark Underwood (to Everyone)**: 2:03 PM: I do worry about maintaining clear crosswalks to cloud taxonomy

**Cavan (to Everyone)**: 2:11 PM: be right back

**David W. Boyd (to Everyone)**: 2:13 PM: A lot of the up front stuff in the V1 documents were added by NIST as part of their publication process.

**David W. Boyd (to Everyone)**: 2:15 PM: The material brought forward from other documents into V7 was brought forward because at one point it was to be a summary document. However, that became difficult to maintain

**Mark Underwood (to Everyone)**: 2:16 PM: Dave - agreed

**Tim Zimmerlin (to Everyone)**: 2:25 PM: CMU's SEI started the CMM back around the 1980s. Organizational maturity addresses and measures "processes" for reproducibility and repeatability.

**Tim Zimmerlin (to Everyone)**: 2:26 PM: Just tossing out the word "maturity" brings along a lot of baggage.

**Tim Zimmerlin (to Everyone)**: 2:28 PM: Program offices want measurable "maturity" that provides accurate and precise predictions of scheduled time and cost and product quality.

**David W. Boyd (to Everyone)**: 2:29 PM: Tim, there are a variety of maturity models in use throughout industry. Watts Humphrey and his bretheren at SEI did a great job of creating a pyramid scheme for software maturity (Note - SEI is a FOR PROFIT Organization). However, our thought process was to provide organizations with some guidelines to see if they were ready to implement big data. This was based on the number of failed big data efforts.

**Tim Zimmerlin (to Everyone)**: 2:32 PM: David, I agree with your subtext. Yourdon, DeMarco, Martin, silver bullets...

**Tim Zimmerlin (to Everyone)**: 2:32 PM: David, that does NOT absolve us of big data system predictive models.

**Ann Racuya-Robbins (to Everyone)**: 2:34 PM: Thank you Russell for taking this forward.

**Tim Zimmerlin (to Everyone)**: 2:41 PM: Wo asks whether Russell will boil the ocean?

**William Miller (to Everyone)**: 2:49 PM: XML Metadata can provide transformation and also extensibility. IT is importatn to align with IETF Internet Standards which was a goal by the W3C.

**Tim Zimmerlin (to Everyone)**: 2:51 PM: William, I agree completely. Compressed XML Schema defined data is efficient and very flexible.

**Tim Zimmerlin (to Everyone)**: 2:51 PM: Russell, you did great!

**Mark Underwood (to Everyone)**: 2:53 PM: Russell- thanks - this is moving along. Remember that it's just a draft and others will help with subsequent crafting

**William Miller (to Everyone)**: 2:54 PM: IEEE SA is also intersted in metadata approaches based on open source.

**Mark Underwood (to Everyone)**: 2:56 PM: Tim - My Q is whether this is new. . . Doesn't it sound like the same old BI data warehouse discipline?

**Mark Underwood (to Everyone)**: 2:56 PM: I.e., the data's big, but the metadata methods may not have changed, some argue

**William Miller (to Everyone)**: 2:57 PM: Good point hard coded applications are not flexible or extensible.

**Mark Underwood (to Everyone)**: 3:00 PM: We would need more time to sort out what is domain-specific vs. what belongs in upper ontologies