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

**NBD-PWD-2016/M0555**

**Source: NBD-PWG**

**Status: Draft**

**Title: Web chat from Meeting of August 23, 2016**

**Chat Log C:\underwood\gov\NIST\Big Data\ChatLog 2016\_08\_23 14\_20.rtf**

**David W. Boyd (Private)**: 1:00 PM: Mark - this may end up being a short meeting since we have gotten no inputs to Vol 6. I think what we will do is bring up the latest Vol 6 and talk through the outline and what we need input on.

**David W. Boyd (Private)**: 1:01 PM: We should also discuss how we can get more inputs/participation.

**David W. Boyd (Private)**: 1:01 PM: Let me know when you want to start.

**Mark Underwood (to David W. Boyd)**: 1:01 PM: Good to go

**Mark Underwood (to David W. Boyd)**: 1:02 PM: Hi Laurie!

**Mark Underwood (to Everyone)**: 1:02 PM: Hi Laurie

**Mark Underwood (to Everyone)**: 1:10 PM: FYI https://en.wikipedia.org/wiki/Federal\_enterprise\_architecture

**cavan (Private)**: 1:16 PM: Mark did you get my text for privacy?

**Mark Underwood (to cavan)**: 1:18 PM: Yep, thx, just ack'd it

**cavan (Private)**: 1:19 PM: thanks

**Russell Reinsch (Private)**: 1:20 PM: Nonsense

**Russell Reinsch (to Everyone)**: 1:21 PM: I also see the subject as an attribute of the data

**Mark Underwood (to Everyone)**: 1:22 PM: An actor in the Unified Modeling Language (UML) "specifies a role played by a user or any other system that interacts with the subject." "An Actor models a type of role played by an entity that interacts with the subject (e.g., by exchanging signals and data), but which is external to the subject."

**Jude Ken-Kwofie (to Everyone)**: 1:22 PM: Is the Data Provider an aggregator data sources?

**Mark Underwood (to Everyone)**: 1:22 PM: Jude - that is one example

**Mark Underwood (to Everyone)**: 1:23 PM: Do we need to have a different role for person than is provided in this UML def?

**Russell Reinsch (to Everyone)**: 1:23 PM: Sounds like she is talking about Consumer

**Mark Underwood (to Everyone)**: 1:25 PM: My intuitioin is that attribute is closer to Ann's aim vs. role

**Jude Ken-Kwofie (to Everyone)**: 1:26 PM: Mark - Thanks

**Mark Underwood (to Everyone)**: 1:28 PM: Trying to use "role" in this way is swimming upstream a bit vs. the usual understanding for role.

**Mark Underwood (to Everyone)**: 1:28 PM: "i.e., folks using UML think they already have this covered (albeit weakly from this group's point of view)

**Mark Underwood (to Everyone)**: 1:35 PM: Perhaps the specific concern is whether data provider as person is different from machine/agent providers; the conventional view is that they are not different.

**Tim Zimmerlin (to Everyone)**: 1:37 PM: Data Provider is an empty placeholder. E.g., the stock brings the baby data.

**Tim Zimmerlin (to Everyone)**: 1:39 PM: In the real world, examples of data providers are the Large Hadron Collider IT system, the Hubble Space Telescope control center, London cctv IT system, Twitter, Web crawlers.

**Mark Underwood (to Everyone)**: 1:40 PM: See https://en.wikipedia.org/wiki/Attribute-Based\_Access\_Control

**Tim Zimmerlin (to Everyone)**: 1:41 PM: Everyone should study the Dublin Core metadata. This is the single best starting point for beginners.

**Tim Zimmerlin (to Everyone)**: 1:42 PM: FITS is an outstanding example for scientific images' metadata.

**cavan (to Everyone)**: 1:43 PM: @Tim & FITS - thanks

**Tim Zimmerlin (to Everyone)**: 1:44 PM: Cavan, for Census an XML schema that defines raw census metadata could combine both data and metadata in each XML document. This works very well.

**Jude Ken-Kwofie (to Everyone)**: 1:44 PM: @Tim - Thanks

**Mark Underwood (to Everyone)**: 1:45 PM: Dave- What seems relevant for today's call is to consider how attribute ("metadata") flows between elements in the RA. Do we think it's handled via orchestration? E.g., how is a role- or attribute set handed off from a provider to a consumer?

**David W. Boyd (to Everyone)**: 1:46 PM: Yeah - I want to see a discussion about providing write ups to the three types of exchanges on the diagram.

**Tim Zimmerlin (to Everyone)**: 1:47 PM: Cavan, each subsequent edit / update could reference the original version or any specific updated version, together with the updated data. Referencing whole documents by URL / URI / URN is space efficient...and avoids spreading many different copies from many different sources...to avoid such confusion.

**Mark Underwood (to Everyone)**: 1:52 PM: See also Audit

**Jude Ken-Kwofie (to Everyone)**: 1:56 PM: is there a testing activity?

**cavan (to Everyone)**: 1:58 PM: @Tim, I agree about somehow referencing the original (almost like a bit chain), about the URL reference, if you are dealing with private data would a URL/URI be available?

**Mark Underwood (to Everyone)**: 2:02 PM: There are some curiosities. ISO 42010 uses stakeholder vs. roles

**Tim Zimmerlin (to Everyone)**: 2:03 PM: Mark, our world is imperfect. This is not a greenfield architecture.

**Mark Underwood (to Everyone)**: 2:05 PM: Tim - I'm trying to rule out all work that we can avoid putting on our plate

**Tim Zimmerlin (to Everyone)**: 2:05 PM: Mark, Frank is going to do this work.

**Mark Underwood (to Everyone)**: 2:08 PM: The misgiving I have here is that the Federal reference (the five threads) have a security model; ISO 42010 does not

**Mark Underwood (to Everyone)**: 2:09 PM: Tim - agree; we have to bridge to the dominant frameworks

**Tim Zimmerlin (to Everyone)**: 2:12 PM: Note UML diagram being used in 42010 document. UML is another big standard from OMG.

**David W. Boyd (to Everyone)**: 2:15 PM: From current 20547-3 draft:

**David W. Boyd (to Everyone)**: 2:15 PM: 3.2.1. reference architecture

a Reference Architecture is an authoritative source of information about a specific subject area that guides and constrains the instantiations of multiple architectures and solutions

[SOURCE: DoD, xxxx]

**David W. Boyd (to Everyone)**: 2:16 PM: NOTE Reference architectures generally serve as a foundation for solution architectures and may also be used for comparison and alignment of instantiations of architectures and solutions

[Editor’s note: need definitive source and need to reach out authors for ISO 42010 for reference architecture. To consider adding other definitions to 42010]

**David W. Boyd (to Everyone)**: 2:17 PM: 3.2.1. role

a set of activities (clause 3.2.1) that serves a common purpose

[SOURCE: ISO/IEC 17789:2014]

3.2.2. sub-role

a subset of the activities (clause 3.2.1) of a given role (clause 3.2.5)

**Frank Farance (to Everyone)**: 2:19 PM: Agree with DoDAF. Yes, when I've built a "reference architecture" is was in reference to "solution architectures" or "inplementation architectures"

**Russell Reinsch (to Everyone)**: 2:20 PM: 2