**NIST BD-PWG Feb. 16 2016 Meeting Notes + Web Chat**

**NBD-PWD-/M0500 + Notes + Web Chat**

**Source:    NBD-PWG**

**Status:     Draft & Append**

**Title:       Meeting Note and Web Chat for Feb. 16, 2016 by Russell Reinsch**

Meeting logistics

Date/time: Feb. 16, 1:00PM – 3:00PM EDT

Web conferencing tool: <https://global.gotomeeting.com/join/790820565>

Audio: Use your microphone and speakers (VoIP) - a headset is recommended, or, call in using your telephone (US, long distance): +1 (646) 749-3122, access code/meeting ID: 790-820-565

## **Agenda**

1. Last week action items
   1. Discuss identified ~6 missing items in Vol. 1 Definition, Nancy and Ann
      1. Whitepaper implication of Big Data in social, business, public, knowledge extraction, and lifecycle, etc.
2. Planning face-to-face meeting at NIST for NBD-PWG, Sept. 20 - 22, 2016 (tentative)
3. **Call for Contributors on NBDIF V2**

* **Continue to seek contributors**
* **Schedule contributors for what areas of document would like to contribute and share/outline their thoughts by given dates**

**[NBDIF V2 Assessment before March 20 Announcement for September F2F Workshop]**

1. From Vol. 1 Definitions, Section 1.5 Future Work (with updates) – Nancy
   1. Enhance existing V1 definitions

Lee Anne – share thoughts/outline what specific contribution by Feb. 23??

Thomas – share thoughts/outline what specific contribution by Feb. 23??

Dan Gillman – share thoughts/outline what specific contribution by Feb. 23??

* + 1. Revisit Big Data and associated definitions
    2. Redo the Data Science section (Nancy feels this has been too diluted in the current version)
    3. Redo data lifecycle to analytics lifecycle or something else
    4. Decide how to address the data-information-knowledge distinctions
  1. Defining the different patterns of communications between Big Data resources to better clarify the different approaches being taken;
     1. Bob Marcus Internet of Things (IoT) (e.g., streaming analytics, other new scenarios)
     2. Bob’s earlier use cases
     3. Application provider with 3V’s (TC69 relations)
  2. Improving the discussions of governance, value, and data ownership;

Tim - shared his outline doc submitted to reflector(?) 2-16: Governance Provenance Curation. Sourced Morningstar. Mark suggests [pillar] section #6 is probably the most important. Next / action item for group - Take contributions on relationship to previous work by NARA, maybe something from Quyen.

Rose/MITRE team – share thoughts/outline what specific contribution by Feb. 23??

* + 1. Governance – architecture view (from velocity, volume, etc.); big data inside/outside ??, Rose
  1. Developing the Management section; (Scope? Obviously systems, etc.)

Russell – share thoughts/outline what specific contribution by March 01.

* 1. Developing the Security and Privacy section; (Arnab, Mark)

Rose/MITRE team – share thoughts/outline what specific contribution by Feb. 23??

* 1. Improve discussion of new behavior (new design patterns?) in big data
     1. Concurrency
     2. Emergent behavior “Matrix Effect” where for example now have PII concerns

Cavan – share thoughts/outline what specific contribution by Feb. 23??

* + 1. Memory innovations

Tim – share thoughts/outline what specific contribution by Feb. 23??

* 1. Whitepaper: Implication of Big Data in social, business, public, knowledge extraction, and lifecycle, etc.

Ann – Share outline today (Feb. 16)??

* + 1. Discuss the topic of value
  1. Architectures (briefly introduce, use GFox descriptions, etc. Brief here, fleshed out elsewhere in Taxonomy. Tim recommends.)
  2. Orchestrator / Orchestration: Refer to additional detail in the MindMap, (a discussion ensued re: Orchestrator as role, but also as conceptual underpinning for enterprise (“business”) process in the RA. (Remarks from Ann, Russell, Bob, Dave, Nancy).
  3. Cross references to other volumes…
  4. INCOSE bib. site

1. From Vol. 2 Taxonomies, Section 1.5 Future Work (with updates) – Nancy
   1. Align with the other v1 documents (Nancy: “not mature” at this stage). *See also* the Mindmap, a resource for this work. [Russell]

Ann – share thoughts/outline what specific contribution by Feb. 23??

Dan Gillman – share thoughts/outline what specific contribution by Feb. 23??

Thomas – share thoughts/outline what specific contribution by Feb. 23??

* 1. (<https://www.mindmeister.com/322462463>)
  2. Translate mindmap into text

Russell – share thoughts/outline what specific contribution by Feb. 23.

* 1. The Subgroup is continuing to explore the changes in both Management and in Security and Privacy. As changes in the activities within these roles are clarified, the taxonomy will be developed further.
     1. The Privacy taxonomy draft is a small fork/task (SnP subgroup)
  2. In addition, a fuller understanding of Big Data and its technologies should consider the interactions between the characteristics of the data and the desired methods in both technique and time window for performance. These characteristics drive the application and the choice of tools to meet system requirements. Investigation of the interfaces between data characteristics and technologies is a continuing task for the NBD-PWG Definitions and Taxonomy Subgroup and the NBD-PWG Reference Architecture Subgroup.
  3. Finally, societal impact issues have not yet been fully explored. There are a number of overarching issues in the implications of Big Data, such as data ownership and data governance, which need more examination. Big Data is a rapidly evolving field, and the initial discussion presented in this volume must be considered a work in progress.
  4. Explore the taxonomy’s ability to work with Geoffrey’s blend of HPC and Big Data
  5. Explore formal methods to define taxonomy (Dave suggested) [Russell] DODAF like?
  6. Efforts should leverage definitions produced elsewhere to avoid ocean-boiling or unresolvable concerns that are less salient for big data. E.g., taxonomy of application patterns (Dave). Emphasize building blocks, roles (Wo).

1. From Vol. 3 Use Cases & Requirements, Section 1.5, Future Work (with updates) – Geoffrey and Piyush

Lee Anne – share thoughts/outline what specific contribution by Feb. 23??

William – share thoughts/outline what specific contribution by Feb. 23??

[Need to come up with a strong motivation reason why people would submit new use cases. It is also hard to get security and privacy use cases]

* 1. Review, finalize, and begin collecting new use cases based from UC Template V2 (Challenge is to foster motivation)

(In parallel: review and identify security and privacy related issues/requirements from Bob’s 10 scenarios and/or the existing 51 use cases)

(Mark is willing to take time to interview people who have BD applications; Mark could Piyush after the call for NASA’s applications)

* 1. How to analyze new use cases with SnP info? How to coordinate vol. 4?
  2. Draw on the use case classification to suggest classes of software models and system architectures [1][2][3][4][5]
  3. A more detailed analysis of reference architecture based on sample codes that are being implemented in a university class. [6]
  4. Collect benchmarks that capture the “essence” of individual use cases.
  5. Additional work may arise from these or other NBD-PWG activities. Other future work may include collection and classification of additional use cases in areas that would benefit from additional entries, such as Government Operations, Commercial, Internet of Things, and Energy. Additional information on current or new use cases may become available, including associated figures. In future use cases, more quantitative specifications could be made, including more precise and uniform recording of data volume. In addition, further requirements analysis can be performed now that the reference architecture is more mature.

1. From Vol. 4 Security & Privacy, Section 1.5, Future Work (with updates) – Arnab and Mark
   1. How to coordinate with vol.3 new use cases on SnP?
   2. Developing the unified security and privacy taxonomy:

William – share thoughts/outline what specific contribution by Feb. 23??

* + 1. Developing the connection between the security and privacy fabric and the NBDRA (should we rename Vol. 4 to “Security and Privacy Fabric”?);
  1. Exploring governance, risk management, data ownership, and valuation with respect to Big Data ecosystem, with a focus on security and privacy;

Rose/MITRE team – share thoughts/outline what specific contribution by Feb. 23??

* 1. Contextualizing the content of Appendix B in the NBDRA; and
  2. Expanding the privacy discussion within the scope of this volume;
     1. Exploring privacy in actionable terms based on frameworks such as those described in NISTIR 8062 [7] with respect to the NBDRA.
  3. Whitepaper: Privacy in Big Data System

Cavan and team – share thoughts/outline what specific contribution by Feb. 23??

Other complement topics (Mark):

1. **V2 “Synthesis” Document**. The scope as proposed is likely beyond the reach of the WG resources. Proposing a first step in lieu of a full V2 as an overview which: (1) summarizes V1 in a more digestible format; (2) addresses conflicts in terminology across the documents; (3) better distinguishes where we to contribute vs. other standards groups; (4) focus on fewer but more detailed use cases (below)
2. **Two primary use cases.** Selecting a big science and a PII-rich use case which can be implemented in the NIST hybrid cloud and highlight Big Data Variety and Velocity. Probably a streaming use case that simulates IoT end points and uses analytics aggressively.
3. **Implementer Guide** that could be adapted by agencies wanting to get their feet wet in Big Data; identify where additional support might be needed (e.g., what a supporting SOW might need to have in it);
4. **Flesh out Orchestration** (6) walkthrough and/or actual example(s) of orchestration across logical boundaries in the RA, probably superimposed on Docker (see also Amazon SWF, Google Kubernetes, Apache Falcon, Spring Cloud Dataflow, Bluemix Websphere Cast Iron, Apcera, Mesos, etc.). Focus on data movement across RA and organizational boundaries. Our examples need to incorporate risk, ownership, provenance, etc. and go beyond cloud orchestration alone by addressing, e.g., sensor data and on-premises applications. Rose/MITRE team – share thoughts/outline what specific contribution by Feb. 23??
5. **SnP Version 2** Two tracks: one Big Data SnP Fabric: Technical, one a narrative: Big Data SnP Fabric: Emerging Processes. The narrative track adapts takes the crosswalk standards document (the long Word doc with a list of standards or recommended practices) and offers a non-proscriptive discussion which references the RA whenever possible. (It can morph to something that supports conformance documents, but that might not happen in the proposed timeframe). The narrative incorporates improvements to taxonomy and makes clearer where policy and technology solutions for SnP can be separated (this is a recurring source of confusion).
6. **Value, Governance, Ownership** There are overlaps with systems management (SysMan), but key aspects are related to Big Data SnP. Create a SysMan section for each major component in the RA, then address value, governance, ownership, privacy etc. in that context. The fabric concept tends to emerge from this level of discussion rather than speaking of privacy or ownership in the abstract.

Rose/MITRE team – share thoughts/outline what specific contribution by Feb. 23??

1. **Big Data Analytics / Data Science** Not mentioned in Wo’s list directly, but the use of “synthetic categories” and classifications through deep learning or other means is a Big Data concern. See this lay discussion on the subject of Big Analytics Failure. It’s related to provenance, but that’s not the whole story.
2. From Vol. 5 White Paper Survey, No Future Work for now
3. From Vol. 6 Reference Architecture, Section 1.5, Future Work (with updates) – David

Thomas – share thoughts/outline what specific contribution by Feb. 23??

Wo – share thoughts/outline what specific contribution by Feb. 23??

* 1. Reference Architecture Refinement
     1. Establish activity and functional component views beyond the current conceptual view
     2. Define high level and general activities and functional components within each view
     3. Identify high level stakeholder and map their concerns to activities and functional components
  2. Reference Architecture application
     1. Establish white paper template (before March 20)

Russell – share thoughts/outline what specific contribution by March 01.

* + 1. Implement the NIST identified six use cases and/or other use cases from the 62 (51 generals and 11 security and privacy) collected use cases or others
    2. Identify development environment (NIST hybrid cloud) for hosting the use case implementations
    3. Create white papers by working with domain experts to identify workflow and interactions among the NBDRA components and fabrics
    4. Review, analyze white papers high-level interactions and workflows and aggregate them into preliminary general interfaces
    5. Conformance approach (emphasized by Frank & Mark)

1. From Vol. 7 Standards Roadmap, Section 1.5, Future Work (with updates) – Russell

Thomas – share thoughts/outline what specific contribution by Feb. 23??

* 1. Examine all version 1 volumes and:
     1. Identify available standards, and those under development. Status: work is in progress.
     2. Gap the differences between version 1 volumes and list of standards, and continue to build and refine the gap analysis and document the findings;
     3. Extend related SDO listing and establish criteria on how to select relevant SDOs
     4. Support harmonizing of terminology between volumes. Status: work has started.
  2. Identify early adopters from academia, government and industry. Status: work has started.
  3. Identify barriers to big data adoption. Doc M0499 uploaded to reflector on Feb 16.
  4. Identify where standards may accelerate the adoption and interoperability of Big Data technologies;
  5. Further map standards to NBDRA components and the interfaces between them.
     1. Enhance gap analysis on how to enable the RA
  6. Document vision and recommendations for future standards activities
  7. Engage communities to attract additional BDWG participation and drive consensus on how big data should move forward. List of whom to engage and email template are 90% complete and first handful of outreach communications went out prior to Feb 23.

[1] Shantenu Jha, Judy Qiu, Andre Luckow, Pradeep Mantha, and Geoffrey C. Fox, “A Tale of Two Data-Intensive Approaches: Applications, Architectures and Infrastructure, in 3rd International IEEE Congress on Big Data Application and Experience Track,” Cornell University Library, June 27- July 2, 2014, <http://arxiv.org/abs/1403.1528>.

[2] Judy Qiu, Shantenu Jha, Andre Luckow, and Geoffrey C. Fox, “Towards HPC-ABDS: An Initial High-Performance Big Data Stack,” Indiana University, August 8, 2014. <http://grids.ucs.indiana.edu/ptliupages/publications/nist-hpc-abds.pdf>.

[3] Geoffrey Fox, Judy Qiu, and Shantenu Jha, “High Performance High Functionality Big Data Software Stack, in Big Data and Extreme-scale Computing (BDEC),” Indiana and Rutgers Universities, 2014. <http://www.exascale.org/bdec/sites/www.exascale.org.bdec/files/whitepapers/fox.pdf>.

[4] Geoffrey C. Fox, Shantenu Jha, Judy Qiu, and Andre Luckow, “Towards an Understanding of Facets and Exemplars of Big Data Applications,” Indiana University, July 20, 2014. <http://grids.ucs.indiana.edu/ptliupages/publications/OgrePaperv9.pdf>.

[5] Geoffrey Fox and Wo Chang, “Big Data Use Cases and Requirements,” Indiana University, August 10, 2014. <http://grids.ucs.indiana.edu/ptliupages/publications/NISTUseCase.pdf>.

[6] Geoffrey Fox. “INFO 590 Indiana University Online Class: Big Data Open Source Software and Projects,” Indiana University, 2014 [accessed December 11, 2014], <http://bigdataopensourceprojects.soic.indiana.edu/>.

[7] DRAFT Privacy Risk Management for Federal Information Systems

<http://csrc.nist.gov/publications/drafts/nistir-8062/nistir_8062_draft.pdf>

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

**NBD-PWD-2015/M0503**

**Source: NBD-PWG**

**Status: Draft**

**Title: Web chat from Meeting of Feb. 16, 2016**

**Chat Log D:\\_wo\1DMG\2015\\_BigDataWG\Docs\ChatLog 2016\_02\_16 15\_20.rtf**

**Russell Reinsch (to Organizer(s) Only)**: 12:56 PM: Using this channel for taking notes NRT.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:05 PM: We is hear das Mutted.....for the benefit of all....Respectfully yours, Pw & Jim did you send me your Rough Drafts....? Just curious as we're in and out of the VA Hospital these days......Respectfullly yours, Pw

**Tim Zimmerlin (to Everyone)**: 1:06 PM: Pw, I sent the draft to DefTax today as an attachment.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:07 PM: Dear Jim: Can we still get a copy...?

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:07 PM: pwc.pwcarey@gmail.com.....

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:09 PM: Dear Jim: We'll even sign an NDA is you require....?

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:10 PM: Rather....'if' you require....

**Tim Zimmerlin (to Everyone)**: 1:10 PM: Pw, pls check your email...

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:11 PM: Jim: Ok...will do...& thanks....

**Russell Reinsch (to Organizer(s) Only)**: 1:11 PM: Eric Harper speaking about tech report on IFC report on industrial internet.

**Tim Zimmerlin (to Everyone)**: 1:11 PM: You should have already got the same attachment today.

**Ann Racuya-Robbins (to Everyone)**: 1:12 PM: I missed what organization Erik is from?

**Ann Racuya-Robbins (to Everyone)**: 1:13 PM: Is it ICC International Code Council?

**Mark Underwood (to Everyone)**: 1:13 PM: Ann - see this http://www.natlawreview.com/article/internet-things-and-inevitable-collision-product-liability-part-5-security-and

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:13 PM: Dear Jim: Looks good....looks quite good, as we're including the US Constituion and the US Bill of Rights within the phases of our Phalanx Fraud Forecasting & Prosecution Too........Thanks again....Pw

**Tim Zimmerlin (to Everyone)**: 1:13 PM: Ann, Industrial Internet Consortium.

**Ann Racuya-Robbins (to Everyone)**: 1:14 PM: Thanks Mark and Tim

**Russell Reinsch (to Organizer(s) Only)**: 1:14 PM: IIC test beds validate his RA.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:20 PM: The future of Big Data and The Cloud and Mobile eco-systems, including Social Media are moving towards individual anonymity...a slow progression....but that's the direction in our personal and professional opinion.....give it ten years....

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:23 PM: Who ever can provide the golden egg for anonimity and accessability...via individual security & privacy....will have a winner....no?

**Mark Underwood (to Everyone)**: 1:24 PM: Orchestration needs to be a prominent part of our test bed

**Russell Reinsch (to Organizer(s) Only)**: 1:25 PM: Ann speaking about implications for life. Frank disagreeing that there is a concern over parallel and concurrency affecting the scientific method. Addressed back in the 60's.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:26 PM: My entire R&D Team likes the way Frank thinks.....

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:27 PM: Dear Ann....Nice/good initial effort.....Respectfully yours, Pw

**Russell Reinsch (to Organizer(s) Only)**: 1:27 PM: Ann: didnt have big data in the 60's. Frank requesting further substantiation.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:28 PM: Now we have something to direct our ray gun brain cells against....like a laser.....thanks Ann....

**Russell Reinsch (to Organizer(s) Only)**: 1:29 PM: PW continues to introduce nuisance jokes into the chat line.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:32 PM: Dear Ann & Frank: Would you both like to take me out to lunch.....nothing extravigant...? Respectfully yours, Pw

**Russell Reinsch (to Organizer(s) Only)**: 1:32 PM: Frank is fully impatient over this as a false argument. Wo eventually referees the discussion.

**William Miller (to Everyone)**: 1:33 PM: It can be said that a eduated person wouyld be able to assure Big Data unfortantely since the data come from the Internet many of the sources are no supplied by reliable sourse who may have applied correct sciencific methods.

**Tim Zimmerlin (to Everyone)**: 1:34 PM: Frank, we all know what you think at this point.

**Tim Zimmerlin (to Everyone)**: 1:35 PM: Repetition does not help make your point.

**Russell Reinsch (to Organizer(s) Only)**: 1:35 PM: Nancy suggests that things actually have changed and we should look at this topic. Frank: these are engineering methods but they do not invalidate the scientific method.

**William Miller (to Everyone)**: 1:35 PM: The wrongnest is obtaining data from non-reliable untrusted souces.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:36 PM: Dear William: Good point, being able to approach with a bit of skepticism and then confirm what is within Big Data is requisite to data provinence...a rush to judgement and/or acceptance serves no one.....

**William Miller (to Everyone)**: 1:36 PM: Collecting data form indivisuls for example who are not relaible or trusted to supply information can introduce erros and false information leading to incorrect conclusions. IF this is applied to machine learning the machine will make a increast decision just faster.

**William Miller (to Everyone)**: 1:37 PM: This is a danger of use of bad Big Data

**Frank Farance (to Everyone)**: 1:37 PM: PW: Would love to take you and Ann out to lunch. I'm in NYC, but maybe we could arrange a 3-way call. Work: +1 212 486 4700, Cell: +1 917 751 2900

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:38 PM: Dear Frank: We should be on the East Coast in a couple of months....we'll keep in touch...Pw

**NANCY GRADY (to Everyone)**: 1:39 PM: @ Frank, hmm I'd have to provide my own lunch since I'm in TN. You going to be at the Dublin ISO meeting?

**William Miller (to Everyone)**: 1:39 PM: you need an advocate or a proxy as Tim suggested

**William Miller (to Everyone)**: 1:39 PM: taking data directly without trust and assurance invites problems

**Tim Zimmerlin (to Everyone)**: 1:40 PM: William, MIT and Princeton already work on your concern. They have generalized algorihms into methods which reduce to algorithms under specific conditions of partial transparency.

**Russell Reinsch (to Organizer(s) Only)**: 1:41 PM: Tim suggests using proxies in cyberspace.

**David Boyd (to Everyone)**: 1:41 PM: So Ann brings up an interesting point. At some point in the future will there be Big Data Haves and Big Data Have Nots and what will be the effect. This could apply to both organizations and individuals.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:41 PM: Dear Tim: We believe in Locards Law....a criminal always leaves trace fingerprints.....(aka: digital fingerprints for our purposes....)....the hard part is finding, and identifying them....but it is possible.....as proven by the Enigma Team identifying specific wireless encrypted messages from specific transmission machines and individual Nazis....

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:42 PM: That's trace fingerpints of forensic evidence....to be sure....

**Cavan (to Everyone)**: 1:44 PM: We are looking into the ability to fingerprint data so that when it is transferred with metadata that we can trust it. Tim if you know any of the researchers working in that area, I would love to talk to them. Thanks

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:44 PM: Dear Tim: Me too....

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:45 PM: Dear Tim: Rather, just point us in the right direction.....

**Lee Anne Davies (to Everyone)**: 1:45 PM: What would the location of the F2F meeting tentative for September?

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:45 PM: Dear Lee Anne Davies: We prefer Holland....

**Cavan (to Everyone)**: 1:46 PM: Micah Altman at MIT, has a digital fingerprint that he did in R for data transfer that was part of the Harvard Dataverse project. I don't know where it stands now.

**Tim Zimmerlin (to Everyone)**: 1:46 PM: Lee Anne, Gaithersburg MD.

**Tim Zimmerlin (to Everyone)**: 1:46 PM: Cavan, Micah sounds like one of the speakers...

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:47 PM: Dear Tim: So Holland is out...? Ok, that's fine by us.....

**Lee Anne Davies (to Everyone)**: 1:47 PM: Ah - I see that it is not decided so the Rocky Mountains in Canada would be lovely to meet!

**Cavan (to Everyone)**: 1:47 PM: Tim, maybe speakers where (I am slow)

**Tim Zimmerlin (to Everyone)**: 1:48 PM: Cavan, I watched the YouTube…based on MIT's TCS group work products…Micah sounds like the post doc / Princeton Assistant Professor.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:49 PM: Dear Anne: You're safer booking your flight to Gaithersburg, MD.....Honest....Pw

**Mark Underwood (to Everyone)**: 1:49 PM: In reviewing Ann's doc on underserved pops, it became clearer to me that Big Data has a particular relationship to mobile data that is less applicable for some other standards groups. It's a concept that needs further development, but starts w/ the greater ubiquity of mobile phones (the new "personal computer") and mobile phones as a primary end point (even for personal IoT considering health wearables).

**Cavan (to Everyone)**: 1:50 PM: Tim, I will look into it. I do think that Micah is at MIT now, but it sounds promising - Thanks

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:50 PM: As usual.....we agree with Mark....

**Tim Zimmerlin (to Everyone)**: 1:51 PM: Cavan, automata theory underlies everything you require; reduction of specific algorithms to general machines; description of sets, complexities, information in methods.

**Tim Zimmerlin (to Everyone)**: 1:52 PM: Cavan, MIT and Princeton assume the input data has been tampered with in invertable ways.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:52 PM: Dear Tim: Do keep your ideas flowing, at your convenience...as we find them quite beneficial....Thanks....Pw

**Mark Underwood (to Everyone)**: 1:52 PM: E.g., by way of illustration, here's a proposal to use phones as "authentication hubs" for smart homes http://bit.ly/1olJvhT

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:53 PM: Dear Mark: Thanks.....

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:54 PM: Dear Tim: My last interruption.....is the US Bill of Rights addressed within your initial effort....?

**Cavan (to Everyone)**: 1:55 PM: Tim, when you get a chance, can you share that url?

**K. Eric Harper (to Everyone)**: 1:57 PM: Very interesting ideas, data ownership is part of this discussion.

**K. Eric Harper (to Everyone)**: 1:58 PM: Unfortunately I have to leave for another meeting.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 1:59 PM: Dear Tim: Q: How can you tell if a Lawyer and/or a Politician is lieing.....? A: They're lips are moving....

**Mark Underwood (to Everyone)**: 2:00 PM: Eric - Thanks for stopping by!

**Russell Reinsch (to Organizer(s) Only)**: 2:00 PM: Tim covering his pillars doc on Gov, Prov, Curation.

**NANCY GRADY (to Everyone)**: 2:02 PM: @Tim - what parts of this have changed with big data? This seems to span all of governance?

**Russell Reinsch (to Organizer(s) Only)**: 2:02 PM: Ann: this doc is organization centric; not user oriented.

**Mark Underwood (to Everyone)**: 2:03 PM: Ann do we have some use case for non-enterprise governance? Are u thnking of advocacy groups or professional associations?

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:03 PM: & According to Anthony Scalia US law comes down to two words....."Who Decides"....and in his opinions it's the people that should establish law, not unelected officials, lawerys and/or judges.....(aka: We the people....)

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:04 PM: oops....rather lawyers.....

**Ann Racuya-Robbins (to Everyone)**: 2:05 PM: Mark we need a non-enterprise governance use case. I have one and will forward.

**Russell Reinsch (to Organizer(s) Only)**: 2:05 PM: I do not like Google Docs.

**Mark Underwood (to Everyone)**: 2:05 PM: It's worth pointing out that Amazon hasn't made money from Big Data in their marketplace/clearinghouse. They have profited from AWS, not their break-even retail ops funded by the stock market

**Mark Underwood (to Everyone)**: 2:05 PM: ebay is a better example

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:05 PM: Dear Ann; Can we also get a copy.....?

**Ann Racuya-Robbins (to Everyone)**: 2:06 PM: sure

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:06 PM: Dear Ann: Thanks.....pwc.pwcarey@gmail.com.....

**Mark Underwood (to Everyone)**: 2:06 PM: Tiim - +1 for centrality of provenance metadata. Spot on.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:08 PM: Dear Mark: Good point....the very same metadata that has been used in the past to direct missle strikes.....Public record....

**Russell Reinsch (to Organizer(s) Only)**: 2:08 PM: XML is a meta language.

**Russell Reinsch (to Organizer(s) Only)**: 2:10 PM: PDFa, zip, JSON not a formal standard from all perspectives. OASIS etc not support it.

**Russell Reinsch (to Organizer(s) Only)**: 2:11 PM: XML works for provenance; can change the name...

**Russell Reinsch (to Organizer(s) Only)**: 2:12 PM: Frank: no semantics, [xml], JSON is digestible, ISO standard. XML vs. JSON.

**William Miller (to Everyone)**: 2:13 PM: json is more for constrained devices

**William Miller (to Everyone)**: 2:13 PM: XML is more robust for semantics

**Cavan (to Everyone)**: 2:14 PM: I am not voting on this, but the Census API moved from XML to jason. Both obviously have reasons to support them.

**Cavan (to Everyone)**: 2:15 PM: Census api for metadata

**Russell Reinsch (to Organizer(s) Only)**: 2:15 PM: Mark: IOT dealing with the sensor edge / section 6.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:16 PM: In our estimation....all data in transit.....is susceptable to Man in The Middle (MTM) spoofing.......and needs to be evaluated and judged via numerous tests for modification; hash tags, scripts, message length, et al......

**Mark Underwood (to Everyone)**: 2:17 PM: The battle over metadata has been fought in the BI / data warehouse space with mixed results. We should learn from that and not try to refight those battles

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:18 PM: Dear Mark: We agree....not to re-fight...but to offer up a nice 'Swiss Army Knife' of tools to assist our audience in conducting their business....no?

**Cavan (to Everyone)**: 2:20 PM: I agree with Mark on being careful about getting bogged down in the swamp of the one and best metadata standard.

**Ann Racuya-Robbins (to Everyone)**: 2:20 PM: Tim: What do you mean by encode?

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:21 PM: Dear Tim: Good effort....have you been looking over my shoulder as we've been building our Phalanx tool....?

**Russell Reinsch (to Organizer(s) Only)**: 2:22 PM: data encodes info. info is a model of a physical system. still machine. knowledge is human.

**Ann Racuya-Robbins (to Everyone)**: 2:25 PM: I would like to offer an alternative definition/description of data,information,knowledge in the big data context because all of these in big data contain bits and shards and sometimes whole pieces of human behavor.

**Mark Underwood (to Everyone)**: 2:27 PM: What relevance to ISO 38500:2015? Overall IT governance. Should help us to peel out the big data piece where we can make a diff

**Ann Racuya-Robbins (to Everyone)**: 2:27 PM: What is raw data then in your view Tim?

**Frank Farance (to Everyone)**: 2:28 PM: 3.4.10

datum (pl. data or datums)

designation whose concept is a value [ISO/IEC 20944-1]

NOTE 1 A datum is created when its designating relationship comes into existence, such as when an instrument or observer records a datum or a computational device emits a datum. Thus, a datum can be thought of as having three parts: a concept, a signifier, and a designating relationship that associates the concept and the signifier.

NOTE 2 Computational devices, largely, are signifier processors not concept processors. The art of software engineering in data processing involves peeling the concepts from the signifiers, yet retaining them nearby via artful and careful programming, and via explicit metadata.

**Frank Farance (to Everyone)**: 2:30 PM: 3.4.11

information

<information science> given context of an object, such as a concept system, that gives it meaning [ISO/IEC 20944-1]

NOTE 1 It is possible to give context using techniques other than concept systems.

NOTE 2 A given context might apply to more than one object, such as a class of objects.

NOTE 3 With respect to data and information, a datum already has meaning: its value (concept). Additional meaning might be provided, such as: revealing one or more concept systems that datums belong to (e.g., relations and relationships among data); describing the circumstances of the act of designation (e.g., who-what-when-where-why-how the data was created or changed); providing mappings to/from the designations, their signifiers, and/or their values (concepts). Other methods are possible for giving additional meaning.

**Mark Underwood (to Everyone)**: 2:30 PM: This was the Open Group's work on IT governance during the SOA era: https://www.opengroup.org/soa/source-book/stds/gov.htm

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:30 PM: Dear Tim: Lying with Big Data....with some aspects of Big Data such as committing Financial Fraud....Human Nature provides us with a great leg up....in tracking, tracing, identifying and documenting, forecasting and prosecuting financial fraud...with Big Data providing one of the mechanisms for same......Thanks Tim....

**Frank Farance (to Everyone)**: 2:31 PM: NOTE 4 It is possible to provide successive contexts of information for data, each revealing more information, the result of each iteration (information) can itself be considered data for the next iteration of revelation, which produce "layers" of information and data. The reverse process is possible, too: each layer of information is stripped of some context that produces data; then the data itself is treated as information and a second iteration of context is stripped from that information to produce data.

**Russell Reinsch (to Organizer(s) Only)**: 2:31 PM: Mis representation of the truth. 1 - 10 are the bad. 11 is the good.

**Frank Farance (to Everyone)**: 2:31 PM: NOTE 5 Because of the successive nature of revelation or stripping, it may appear that terms data and information can be used interchangeably, but this is incorrect. Data is characterized by the signifiers, their associations with concepts, and their notions of equality; information is characterized by referencing the context(s) overlaid upon the data; and both might be present. Likewise, because context can always be added or stripped, it is impossible to say that something is purely data (but no corresponding information) or purely information (but no corresponding data).

EXAMPLE 1 One kind of context is defining the symbols used for communication, e.g., an encoding.

EXAMPLE 2 If X is a sequence of bits that represents an encrypted message (the object) whose meaning is merely a sequence of designations whose symbols are zero and one; then the decryption key is an example of context (a mapping) that gives additional meaning to the data to reveal information (i.e., the decoded message).

**Frank Farance (to Everyone)**: 2:31 PM: EXAMPLE 3 The designation "20" is a datum, a kind of object. The context "temperature in degrees Celsius of New York City at 2003-07-19 16:00 UTC" is information about this object.

**Mark Underwood (to Everyone)**: 2:32 PM: Feigenbaum

**Mark Underwood (to Everyone)**: 2:34 PM: Tim, yeah I guess this issue of long term preservation is a Big Data issue. Hadn't thought of that. Drexel folks have some contracts to stsudy that problem.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:34 PM: Dear Mark & Tim: Does this also come down to....."Who Decides".....?

**Russell Reinsch (to Organizer(s) Only)**: 2:35 PM: Tim doc is a big roast pig to pick from.

**Russell Reinsch (to Organizer(s) Only)**: 2:38 PM: Frank: bit rot has been addressed by NARA, etc.

**Mark Underwood (to Everyone)**: 2:39 PM: FYI Quyen Nguyen on today's call

**Tim Zimmerlin (to Everyone)**: 2:44 PM: Mitre could do a very good job here.

**Russell Reinsch (to Organizer(s) Only)**: 2:44 PM: Nancy concerned w scope. Encourage more thought to be only about big data.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:45 PM: Sure: We'll just tell our audience that the following areas are worthy of addressing, but not at this point in time.........

**Mark Underwood (to Everyone)**: 2:45 PM: +1 Nancy- I was going to suggest we need to have part of the governance conversation address how systems management in the RA can be used in BD governance - as different from non BD systems, using the RA as a lens

**Mark Underwood (to Everyone)**: 2:46 PM: - Even tho Tim has reservations about the RA in its current state; it would still help guide others

**Russell Reinsch (to Organizer(s) Only)**: 2:49 PM: Frank: look at devops and small shops.

**Mark Underwood (to Everyone)**: 2:50 PM: A good use case would be a compliance scenario that requires BD scale

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:51 PM: Dear Mark: Checkout Pw's Financial Industry Use Case......we hear it's a pip......Hope this helps....Pw

**Russell Reinsch (to Organizer(s) Only)**: 2:52 PM: Tim: Ann and PW believe governance is new. We are subject to the second system paradigm.

**Mark Underwood (to Everyone)**: 2:53 PM: PW- agreed - have you turned up one that's gone so far as to employ Spark or HDFS?

**Russell Reinsch (to Organizer(s) Only)**: 2:54 PM: To Wo: I was going to a refresher from you on what you wanted to see in the whitepaper template.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:54 PM: Dear Mark: Actually, yes....we incorporated Hadoop File Structure within our Use Case....along with several SEC, PCAOB requisit regulations......both HW and SW.....

**Russell Reinsch (to Everyone)**: 2:55 PM: I have to step out for 2 minutes

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:55 PM: oops 'requisite' instead...

**Mark Underwood (to Everyone)**: 2:55 PM: PW would be great to interview somebody in finance using HDFS etc - or forensics for that matter.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:56 PM: Dear

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:56 PM: Dear Mark: That's us......

**Mark Underwood (to Everyone)**: 2:56 PM: Ah

**David Boyd (to Everyone)**: 2:56 PM: Wo I have to go. I have a 3PM to attend.

**David Boyd (to Everyone)**: 2:57 PM: Please see my email to the reflector on the request for contributions.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 2:57 PM: Yep....we' re focuing on forecasting and prosecuting Financial Fraud...via our tool, in development.....at this point in time....

**Mark Underwood (to Everyone)**: 2:59 PM: A useful aspect to Governance in the Open Group doc is artifacts, such as service descriptions. This could include declarations of privacy risk etc

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 3:00 PM: Based upon reams of research....CEOs prefer intuition, experience and hunches when making their minds up.....Go figure....eh?

**Mark Underwood (to Everyone)**: 3:00 PM: Tim, did you call me out as an Alpha!

**Tim Zimmerlin (to Everyone)**: 3:00 PM: Yep!

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 3:00 PM: Dear Mark; You must be thinking of Jack London.....no?

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 3:02 PM: Dear Sir: Another good meeting.....Thank you......We'll help anyway we can......Respectfully yours, Pw

**Mark Underwood (to Everyone)**: 3:02 PM: 'PW <Howling> Note angle brackets.

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 3:02 PM: Dear Mark: Thank you....it's hard to find a decent audence these days.....don't you agree....?

**Ann Racuya-Robbins (to Everyone)**: 3:06 PM: got to go thank you all

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 3:06 PM: Dear Ann: Drive safely.....

**Russell Reinsch (to Everyone)**: 3:09 PM: Tim recommend ignite. xsee,

**Wo Chang (to Everyone)**: 3:09 PM: thanks Russell

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 3:11 PM: Lets not forget 3-D Printers producing body parts for animals, that apparently work.....go figure....

**Pw Carey, Compliance Partners, LLC (to Everyone)**: 3:12 PM: Dear Sir: Thanks again.....

**Russell Reinsch (to Organizer(s) Only)**: 3:14 PM: ntcoe doc to capture endpoint device could be a starting point for whitepaper template.