NIST Big Data WG Meeting 18-FEB-2014

Prepared by Mark Underwood, Krypton Brothers

# Conferences and Publications

ACM accepted the NIST conference publication offer. Registration deadline and abstracts are due 11 March – 300 words. Full paper is due in June 30. The peer-reviewed papers may be published in ACM’s the international conference series. Wo strongly encourages people to participate.

URL: <https://rd-alliance.org/1st-meeting-notice-isoiec-jtc-1-study-group-big-data.html>

Wo also mentioned the Oct. IEEE 2014 meeting in Washington DC. Wo is co-chair on the industry and government side for this conference. He is looking program committee members and papers from industry and government.

Prof. Xiaohua Tony Hu is general chair for the IEEE conference – (via Drexel). He might have some students to contribute to our RDA-BDI-WG/NBD-PWG works.

William suggests a mini-meeting for D.C. local members around the March 4-5 NIST Data Science Analytics symposium meeting. Wo is to be notified if anyone wants to participate. Tim Zimmerlin and William are likely attendees, and Kori McFadgion (Cisco – welcome!) is a possible from Maryland ([kmcfadgi@cisco.com](mailto:kmcfadgi@cisco.com)).

URL: <http://www.nist.gov/itl/iad/data-science-symposium-2014.cfm>.

Sanjay Mishra mentions another conference involving the UK Research Council and others in March.

INCITS – Keith is the chair for the Big Data AHG for this group. He may plan telecons before the March meeting. This group will provide input from the U.S. interests to the JTC-1 worldwide body for Big Data possible standardization, etc. The San Diego meeting is the first international meeting, where he may participate further.

<http://www.incits.org/dotAsset/6644d0c1-2ad1-4061-9269-5ecf1b75780d.htm>

# Use Cases Discussion

Geoffrey: Proposes that we put all the use cases in one place on the use cases web page to avoid confusion.

William Miller is working on the cargo shipping use case from FedEx et al. He has contacted their Custom Critical system POC, and others by DHL and UPS. FedEx started in 2008 to integrate a sensor service. It’s used for alerts, advisories across domains, though they’re not currently doing that – for good or for bad. William believes there may be an opportunity for third parties to create datasets that could be used by the main carriers given that the state of their current systems. Williams advises that this will likely be unstructured data that is outside of the carriers’ domain.

David Boyd (Data Tactics): One possible shipping oriented data set is the [Automated Identification System (AIS)](http://en.wikipedia.org/wiki/Automatic_Identification_System) which tracks shipping worldwide and is publicly available.

Wo: From the RDA Big Data Analytics Interest Group, they have a few use cases that Wo has discussed with them as they have accompanying public data. These will be added to the other use cases on the consolidated list.

Tim Zimmerlin: There may be some ETL and cleansing data rework needed to make the contributed data usable. Several on the call responded in agreement. Wo mentions that the real world domain knowledge is needed to address this. David Boyd observes that there are typically problematic “artifacts” of such real world systems . Wo concurred, and mentions fusion-related issues as a real-world complexity as well.

Wo – Responding to standards for interacting components -- Suggests we avoid issues of data standardization as distractions from the main intention. At some point generic interfaces could be designed that use the architecture components – but later in the development of standards and practices.

Tim suggested XML/RDF standards for component-component communication, for which JAXB and JAXP are extant parsing frameworks.

Geoffrey Fox uploaded an analysis of usability of 51 use cases in implementation phase:

http://bigdatawg.nist.gov/\_uploadfiles/M0291\_v1\_7418211592.pdf

# Security and Privacy Subgroup

Arnab reviewed the current XLS draft produced by the S&P subgroup. He solicited the group to contribute and review. The CSA group is not using use cases but rather topic-based. Wo says we may have to put some toy S&P frameworks around them.

The S&P XLS document is available for editing at:

<https://docs.google.com/spreadsheet/ccc?key=0AqJGdUJOefindHl0ZHRDZ3VJdDhDREduUjh5d1QzV1E&usp=sharing#gid=0>

Anonymous editing is supported, but logged in updates are preferred to facilitate editing.

Pw Carey (Compliance Partners) offered to submit a Use Case for a Cloud/Big Data Eco-system Digital Forensics/Fraud Detection & Analysis tool.

Atul (IMC ) : Offers a use case for fraud detection and prevention is available. He’s not using traditional Hadoop style methods, but is using viz tools from IBM to prevent supply chain fraud.

Tim suggests that having a requirements traceability matrix would be helpful for CIO-level participants. Suggests that links to products and concepts that are domain-specific (e.g, SoX or HIPAA) would be helpful for prospective readers. A version of this table fleshed out for participate use cases is one approach Arnab suggests. Tim says that possibly a Phase 2 or 3 by NIST could provide still more detail.

# Action Items

1. Place all the use cases in one place on the use cases web page to avoid confusion. Geoffrey
2. Make SnP use cases available from the SnP Requirements as part of the item #1. Arnab
3. Make fraud detection and prevention use case available. Atul

# About Participants

Arnab will host next week’s meeting as Wo is traveling but Dan from NIST will be available.

Orit Levin (Microsoft) is out for the next two weeks.

**WELCOME** New on today’s call: Kori McFadgion (Cisco, [kmcfadgi@cisco.com](mailto:kmcfadgi@cisco.com)) and Atul Mathur ([amathur@imc.com](mailto:amathur@imc.com)).

# Attendees

1. Kori McFadgion
2. Wo Chang
3. Nancy Grady
4. Mike McGee
5. Mark Underwood
6. Geoffrey Fox
7. Davdi Boyd
8. Arnab Roy
9. Manoj
10. Marcia Mangold
11. Orit Levin
12. Atul M
13. Bob Marcus
14. Sanjay Mishra
15. Tim Zimmerlin
16. Pw Carey
17. Dan Samarov
18. William Miller
19. Steve K.

(12:54 PM) Kori McFadgion joined.

(12:55 PM) Kori McFadgion: Good afternoon Wo, thank you for inviting me

(12:55 PM) Wo Chang (Host, NIST): Hi Kori, thanks for joining us!

(12:56 PM) Kori McFadgion: You are very welcome.  I apologize for missing the first couple meetings you held.

(12:57 PM) Nancy Grady (SAIC) joined.

(12:57 PM) Mike McGee joined.

(12:57 PM) Mike McGee disconnected.

(12:58 PM) Mark Underwood (Krypton Brothers) joined.

(12:58 PM) Geoffrey Fox joined.

(1:00 PM) David Boyd (Data Tactics) joined.

(1:00 PM) Wo (guest) joined.

(1:00 PM) Arnab Roy (Fujitsu) joined.

(1:02 PM) Manoj (CyberIQ) joined.

(1:03 PM) Marcia Mangold joined.

(1:03 PM) Orit Levin (Microsoft) joined.

(1:04 PM) Kori McFadgion: I will dial into the bridge

(1:04 PM) Kori McFadgion: thank you

(1:04 PM) Wo Chang (Host, NIST): Phone: 206-402-0823, Participant code: 272-30-504

(1:07 PM) Atul M joined.

(1:07 PM) Bob Marcus joined.

(1:11 PM) Sanjay Mishra joined.

(1:13 PM) Tim Zimmerlin joined.

(1:15 PM) Pw Carey, Compliance Partners, LLC joined.

(1:16 PM) Tim Zimmerlin disconnected.

(1:16 PM) Dan Samarov joined.

(1:17 PM) Tim Zimmerlin joined.

(1:19 PM) Pw Carey, Compliance Partners, LLC: Good afternoon.....we'd like to submit a Use Case for a Cloud/Big Data Eco-system Digital Forensics/Fraud Detection & Analysis tool.....

(1:20 PM) William Miller joined.

(1:21 PM) Geoffrey Fox: To Pw Carey; submit yours-- we can certainly add

(1:23 PM) Pw Carey, Compliance Partners, LLC: Ok...thanks.....

(1:25 PM) Pw Carey, Compliance Partners, LLC: Thank you Marci for that very good Power Point Presentation.....we went through it a couple of times......quite good....also, do you see any need for giving more attention to mobile platforms/footprints.....?

(1:27 PM) Pw Carey, Compliance Partners, LLC: We agree, quite a useful Use Case......

(1:28 PM) Pw Carey, Compliance Partners, LLC: Perhaps FedEx would like to Collaborate with William....?

(1:28 PM) Pw Carey, Compliance Partners, LLC: Regarding providing 'data sets' 'data points'....et al....

(1:29 PM) David Boyd (Data Tactics): One possible shipping oriented data set is the Automated Identification System (AIS) tracks ships worldwide and is publically available.

(1:30 PM) Tim Zimmerlin: I am going on 3rd & 4th.

(1:30 PM) Tim Zimmerlin: Wo, you should brief this conference!

(1:30 PM) Pw Carey, Compliance Partners, LLC: Good point....use as many publically available data sets is advisable....

(1:32 PM) Atul M: My company is involved with some projects on Fraud Prevention. Is this use case of interest here?

(1:32 PM) Pw Carey, Compliance Partners, LLC: Big Data Analytics Symposium....on the 3rd & 4th...will that be viewed via the Internet...?

(1:32 PM) Kori McFadgion: I am local to MD as well

(1:33 PM) Pw Carey, Compliance Partners, LLC: Also, the Cloud Security, Privacy and Forensic WGs are haveing confernce towards the end of March....24th, 25th time frame....

(1:35 PM) William Miller: http://www.nist.gov/itl/iad/data-science-symposium-2014.cfm

(1:35 PM) William Miller: The data science symposium is on March 4,5

(1:36 PM) Pw Carey, Compliance Partners, LLC: Thanks....

(1:37 PM) Marcia Mangold disconnected.

(1:38 PM) Steve K joined.

(1:47 PM) Wo (guest) disconnected.

(1:48 PM) Pw Carey, Compliance Partners, LLC: Are there going to be Demos for the specific Use Cases....?

(1:50 PM) Wo Chang (Host, NIST): Bob: any insight or comment from you for the overall approach?

(1:50 PM) Pw Carey, Compliance Partners, LLC: Demonstrations and/or Power Point Presentations.....of the Use Cases....?

(1:52 PM) Geoffrey Fox: I need to be away 1.55-2.20. I did at http://bigdatawg.nist.gov/\_uploadfiles/M0291\_v1\_7418211592.pdf upload an analysis of usability of 51 use cases in implementation phase

(1:52 PM) Tim Zimmerlin: Communications could be restricted to XML and RDF together with schema.

(1:52 PM) Pw Carey, Compliance Partners, LLC disconnected.

(1:52 PM) Tim Zimmerlin: JAXB and JAXP are ready made Java frameworks to parse these.

(1:53 PM) Geoffrey Fox: http://bigdatawg.nist.gov/\_uploadfiles/M0277\_v3\_4193655353.pptx has one and a bit more slides on each of 51 use cases

(1:55 PM) William Miller disconnected.

(1:55 PM) Pw Carey, Compliance Partners, LLC joined.

(1:56 PM) Pw Carey, Compliance Partners, LLC: My apologies...our network server just re-booted......we just was wondering what will be the format/logistics for the presentations.....?

(1:56 PM) Mark Underwood (Krypton Brothers): he's daying "fraud"

(1:56 PM) Mark Underwood (Krypton Brothers): saying

(1:57 PM) Atul M: Fraud Prevention and detection

(1:57 PM) Pw Carey, Compliance Partners, LLC: We can chat via email as well......

(1:59 PM) Atul M: amathur@imc.com

(2:00 PM) Kori McFadgion: I have to drop off, but please include me in any emails or meeting minutes.  Thank you again! kmcfadgi@cisco.com

(2:00 PM) Kori McFadgion disconnected.

(2:01 PM) Steve K disconnected.

(2:02 PM) Arnab Roy (Fujitsu): https://docs.google.com/spreadsheet/ccc?key=0AqJGdUJOefindHl0ZHRDZ3VJdDhDREduUjh5d1QzV1E&usp=sharing#gid=0

(2:02 PM) Bob Marcus disconnected.

(2:03 PM) Pw Carey, Compliance Partners, LLC: What was the document's number....?

(2:03 PM) Mark Underwood (Krypton Brothers): M0283

(2:03 PM) Pw Carey, Compliance Partners, LLC: Ok...got it....M0283 V1

(2:03 PM) Pw Carey, Compliance Partners, LLC: Thanks...

(2:07 PM) Pw Carey, Compliance Partners, LLC disconnected.

(2:11 PM) Mark Underwood (Krypton Brothers): The document is available for anonymous editing, but it's helpful to know who's contributed so we can edit it properly

(2:12 PM) Pw Carey, Compliance Partners, LLC joined.

(2:13 PM) Nancy Grady (SAIC) disconnected.

(2:13 PM) Pw Carey, Compliance Partners, LLC: Dear Tim: You're breaking up....but send me an email and we'll see if we can help you out a bit...pwc.pwcarey@gmail.com

(2:18 PM) Orit Levin (Microsoft): Good!

(2:19 PM) Orit Levin (Microsoft): I will be out for the next two weeks....

(2:20 PM) Pw Carey, Compliance Partners, LLC: Our sound system is....K A P U T ....

(2:20 PM) Orit Levin (Microsoft): I will send an email as well...

(2:21 PM) Pw Carey, Compliance Partners, LLC: Have a nice time....

(2:23 PM) Geoffrey Fox disconnected.

(2:24 PM) Tim Zimmerlin disconnected.

(2:24 PM) Atul M disconnected.

(2:24 PM) Mark Underwood (Krypton Brothers) disconnected.

(2:24 PM) David Boyd (Data Tactics) disconnected.

(2:24 PM) Orit Levin (Microsoft) disconnected.