## Introduction

This is the first meeting of the NIST Big Data Working Group. The meeting ran from 10AM EDT to 12AM EDT. The meeting ran into a number of communication obstacles which we overcame. Of the minutes captured (78 lines), my rough estimate is 36% were connect/disconnect messages and 11% were mute comments. Some of our attendees use only telephone to join us.

## Attendees[[1]](#footnote-1)

|  |  |  |
| --- | --- | --- |
|  | **Name** | **Comment** |
| 1 | Alicia Zuniga-Alvarado/AZA |  |
| 2 | Bill Mandrick |  |
| 3 | Chris Hawkinson | May be Chris Hawkinson |
| 4 | Christine Hawkinson | May be Christine Hawkinson |
| 5 | Dan Samarov |  |
| 6 | David Boyd |  |
| 7 | Deborah Blackstock (MITRE) |  |
| 8 | Eugene Luster (DISA CTO/R2AD) |  |
| 9 | Gary Mazzaferro |  |
| 10 | Glenda Hayes (MITRE/Army HQ) |  |
| 11 | Karen | Need last name |
| 12 | Nancy Grady (SAIC) |  |
| 13 | Orit Levin (Microsoft) |  |
| 14 | PavithraKenjige11 |  |
| 15 | Pw Carey, (Compliance Partners, LLC) |  |
| 16 | Rod Peterson (Dept of VA) |  |
| 17 | Willaim Miller (MaCT USA) |  |

## Highlighted Conversation Topics/Open Questions to be addressed

1. How will this effort interface with the efforts to create a new web domain.....named: dot secure....?

2. Are we going to incorporate the tools and skill sets associated with Digital Forensics within a Cloud Ecosystem.....?

3. IEEE Big Data conference http://www.ischool.drexel.edu/bigdata/bigdata2013/topics.htm

4. Will we be addressing structured versus unstructured data (this is important related to use with analytic) ?

5. Can we review current work and scope what we need to do for our immediate products that will benefit the government acquisition process?

6. Will we be discussing central versus decentralized systems?

7. Do we need to delineate the idea of transformation from unstructured to structured? I tend to think of that as a continuing versus a clean line. Eg. a spreadsheet is primarily structured but may have unstructured fields.

8. How can we use Big Data to conduct Audits within the Cloud Ecosystem would be a nice outcome of these efforts.....No...? (Possibly Cloud Forensics WG oriented)

9. What about velocity? (The assertion by question writer: it depends on the bandwidth of the network velocity depends on infrastructure”) Velocity can depend on multiple structures....including Access, Integrity, Availability, et cetera.....

10. Will we considering the use of "Veracity" ? Are there Big Data concerns or approaches to dealing with data quality of Big Data. WRT Veracity - There may be different techniques that must be deployed to deal with data quality

11. How does any of our these definitions differentiate big data from traditional data definitions?

12. Will we be using this definition from a recent NSF?

13. If the techniques for dealing with a "V" for Big Data v. non-Big Data, it should be included. Correct?

14. The phrase "big data" in this solicitation refers to large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future.

15. There are some use cases, like longitudinal studies of health outcomes, where a longer lifetime of the data is important. There are longitudinal uses of the data.

16. Will we be considering predictive and prescriptive analysis with big data?

17. Providence is part of the metadata.

18. How clean does data need to be with big data? Because data may not be clean, conflict-resolution techniques are of greater importance.

19. I agree, getting too deep into the specifics of the analytics would not help in specifying taxonomy

20. I think with regards to analytics, the challenge is to be able to make sense out of extremely large Volumes of data, and that change frequently over time. Not unique to big data; more a quality of scale.

21. Will this effort impact the Cloud Computing Risk Intelligence Map?

22. Regarding Big Data security, there is a parallel effort at NIST related to cyber security. Will we might incorporate that work by reference?

23. How do we establish categories for the taxonomy to be sure all is covered?

Physical Objects, Processes, Information bearing entities, information content entities, abstract things, relations between objects/processes

24. How do we measure “scale” for saying it’s “Big Data”? How do we say “It depends”?

25. What needs to be in the taxonomy to help procurement?

26. What are our stakeholders?

27. What are our use cases?

28. Do we want to create an OWL file for the taxonomy?

29. Do we want a two column table for examples of traditional and big data realms?

30. How do we differentiate “big data” from current powerful transaction systems?

31. How do we describe horizontal vs vertical scaling?

32. What security concepts do we need?

33. What open data initiative concepts do we need?

Actions:

Review the subgroup charter for needed enhancements

Consider what has not yet been included in the above lists of concepts

Share existing definitions/taxonomies we can consider

1. This list is not comprehensive. Not all comments are captured from the NIST collaborative tool. [↑](#footnote-ref-1)