**NIST Big Data Working Group (NBD-WD)**

**NBD-WD-2013/M0063**

|  |  |
| --- | --- |
| **Source:** | Reference Architecture Subgroup |
| **Status:** | Draft |
| **Title:** | Reference Architecture Subgroup Minutes for July 25, 2013 |
| **Date:** | July 25, 2013 |
| **Author(s):** | Orit Levin (Microsoft), James Ketner (AT&T), Don Krapohl (Augmented Intelligence) |

**Note taker: Sanjay Mishra (Verizon)**

Topic: Big Data Architecture Framework (BDAF), an outcome of brainstorming session at the University of Amsterdam

Presenter: Yuri Demchenko

Big Data Architecture Framework under the proposed context:

1. Data Models, Structure, Types
2. Big Data Management
3. Big Data Analytics and Tools
4. Big Data Infrastrucutre
5. Big Data Security

Big Data Definition - 5Vs of Big Data

1. Volume
2. Velocity
3. Value
4. Veracity
5. Variety

Big Data Origins

* Science
* Telecom
* Industry
* Business
* Living Environment, Cities
* Social media and networks
* Healthcare

Big Data Target Use

* Scientific discovery
* New Technologies
* Manufacturing, processes, transport
* Personal services campaigns
* Living environment support
* Healthcare support

**Q&A**

**Wo:** we talk similarly in Taxonomy, should we build in the SDL concept in the framework?

**Yuri:** It could be built-in or it could be separate

**Wo:** Now you have those components defined, what kind of interface

**Yuri:** We can define interface easily. Gary has referred earlier to define interface. The technical part is easy but how to define is complex

**Wo:** What kind of interface you may have mind?

**Yuri:** Different layers may have different, for e.g. REST or SOAP based

**Wo:** How could we define actors, roles and responsibilities?

**Yuri:** You must define actors, roles and responsibilities at the earliest.

**Nancy:** How do we define roles? Can we separate out roles at different stages of lifecycle?

**Yuri:** The roles do not relate to lifecycle but depends on Operations (provisioning, management, CRM). The roles are defined in Tele Management Forum document.

**Wo:** Within the life cycle management, is there any specific data you want to include?

**Yuri:** Tele Management Forum may be where this develops.

* Wo will send what he has from ISO.

**Summary:**

Roles view

* ISO
* BD-specific extensions

Data flow view vs work-flow specific view (big data)

Service view

* Language specification

Interface view – data characteristics

**General comments:**

**Gary:** The biggest differentiator for big data with other traditional or commercial applications is that BD is driven by work-flow architecture and is not a data driven model as in the traditional or commercial applications.

**Wo:** Gary, you mention about capability. Can you define in terms of architecture?

**Gary:** Term is overused terminology….Capability and Capacity, I meant in the sense of outcome. What are the outcomes, the goals are.

**Wo:** Are you saying each layer is a different capability? Such as visualization?

**Gary:** Yes. Visualization, you could you any dimensions (two, three or four dimensions). E.g., Take a PDA, you may end up displaying for e.g. a 3-D into a 2-D based on the “visual” capability of the PDA.

**Don:** Do we have a Technical Writer for this group?

*(No resolution to above)*

**Wo:** If you think of mark-up language. If you look at audio, video streaming with programming information….You have different layers….Likewise, Big data may have many different format…There may be different layers….Would this data centric model work here?

**Yuri:** Streaming media format might be a good starting point, I would be interested to investigate.

**Gary :** Look at Open Grid Format.

http://www.ogf.org/dfdl/

**Gary & Nancy:** By the end of Sep, will not be a complete document but rather a progression. Concept should be clearly presented for experts to review and provide their feedback. By Sep 27, we should have concepts defined in terms of Taxonomy and Architecture.

**Next Monday**

A combined meeting with the Taxonomy group. The Architecture group should create some presentations with actors, responsibilities and interfaces…..

**Comments on the Public chat:**

William Miller (MaCT USA): check out this ISO standard that is in progress

William Miller (MaCT USA): it will allow identification of the data, tacking, and is useful in a push. The identification is the key. This will be an international standard

<http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=45240>

Babak Jahromi (Microsoft): Data Controller vs. Data Processsor

Babak Jahromi (Microsoft): former owns/controls/sources the data, and the latter processes/acts on the data. Cloud Providers are often Data Processors

Demchenko (UvA): agree with Gary's slide 34, can also add that BD will also develop in parallel with few quite related technologies and have cross-fertilisation

**Attendees:**

Alicia Zuniga-Alvarado/AZA

Babak Jahromi (Microsoft)

Bob Marcus

Don Krapohl

Gary Mazzaferro

Marge Cole SGT

Nancy Grady (SAIC)

Pratik Thakkar

Quyen Nguyen(NARA)

Sanjay Mishra(Verizon)

Scott Brim (Internet2)

Tim Zimmerlin (Automation Technologies)

Vivek Navale

William Miller (MaCT USA)

Wo Chang (Host, NIST)

Yuri Demchenko (UvA)