**Requirements Working Group Agenda followed by Minutes 9 July 2013**

**Agenda**

**Discuss charter and NIST guidelines**

**Brainstorm about work of Requirements WG**

What to do and how to do it?

**What to do**

**Collect Use Cases**

Specific examples

Generalization to similar use cases

Need to define format of a use case so can generate abstractions

**Use cases linked to abstractions/generalizations of features as below**

Refine abstractions and link to use cases

Network, Storage, Compute needs

Infrastructure/Architecture Requirement (NoSQL v SQL, MapReduce v MPI, Clouds v HPC...)

Data size v Compute Size

Data centralized or distributed

See Bob Marcus document for more examples

Algorithms needed; Implementation needed (e.g. parallel or GPU versions)

Security&Privacy

Training or Expertise needed

**How to do it or Process to follow**

Gather documents and URL's/Citations

Agree on approach; formats & methodology of giving input

Specify in WG meeting

Edit document on Google docs

Upload to <http://bigdatawg.nist.gov/show_InputDoc2.php>

Interactions with other WG

Contact other individuals and organizations that can contribute use cases etc.

**Minutes**

We reviewed the NIST guidelines and discussed the proposed agenda items. The general plan was agreed but a detailed plan to go forward was not fully formulated. The chat text below gives details on discussion of use cases (at what level should they be presented (technology -- user perspective -- application perspective) and the importance of agreeing on a template. We also discussed collaboration technology to use (Wiki v Google docs)

**Chat Record**

(10:59 AM) Karen Guertler: Karen Guertler

(10:59 AM) Karen Guertler: I'll be on mute; however the audio is coming thru loud & clear via my Mac.

(11:00 AM) Geoffrey Fox joined.

(11:01 AM) Orit Levin (Microsoft) joined.

(11:01 AM) Tim Zimmerlin (Automation Technologies) joined.

(11:01 AM) Bob Marcus joined.

(11:02 AM) Arnab Roy (Fujitsu) joined.

(11:04 AM) Bob Marcus disconnected.

(11:05 AM) Bob Marcus joined.

(11:05 AM) Karen Guertler: Today, I can hear via the Web.

(11:05 AM) Karen Guertler: No worries. ;)

(11:06 AM) Orit Levin (Microsoft): He is correct. No audio FROM web.

(11:06 AM) Yuri Demchenko (UvA) joined.

(11:06 AM) Yuri Demchenko (UvA) disconnected.

(11:07 AM) Yuri Demchenko (UvA) joined.

(11:07 AM) Yuri Demchenko (UvA) disconnected.

(11:07 AM) Yuri Demchenko (UvA) joined.

(11:08 AM) Bob Marcus disconnected.

(11:10 AM) William Miller joined.

(11:11 AM) Karen Guertler: Perhaps organize the google docs space to include a category for user profiles.

(11:12 AM) Karen Guertler: If that's possible.

(11:12 AM) Karen Guertler: Great!

(11:13 AM) Bob Marcus joined.

(11:17 AM) Karen Guertler: I do have a few questions; perhaps better answered by posting to the collaborative space.

(11:18 AM) Karen Guertler: For example, I'd like to know whether this initiative will involve a formal RFI.

(11:18 AM) Karen Guertler: to obtain input from the 'universe' of potential stakeholders.

(11:19 AM) Karen Guertler: Sounds good.

(11:21 AM) Bob Marcus disconnected.

(11:23 AM) Bob Marcus joined.

(11:23 AM) Karen Guertler: I can provide a sample use case format; however NIST might have a preferred format.

(11:24 AM) Dusty Jackson joined.

(11:26 AM) William Miller: I would like to contibute a use cases related to chargo shipping

(11:27 AM) Tim Zimmerlin (Automation Technologies): IMHO, this subgroup needs a wiki.

(11:27 AM) William Miller: it is the largest are for use of Big Data

(11:27 AM) William Miller: correction cargo shipping

(11:27 AM) William Miller: sorry for typo

(11:27 AM) Karen Guertler: @ Tim, yes, we've discussed various collaboration options.

(11:28 AM) Tim Zimmerlin (Automation Technologies): My point is we are destined to be overwhelmed by uncoordinated inputs.

(11:28 AM) Tim Zimmerlin (Automation Technologies): Ok!

(11:28 AM) Karen Guertler: agree wrt need for version control.

(11:29 AM) Tim Zimmerlin (Automation Technologies): Ok!

(11:30 AM) Karen Guertler: I'd also like to know whether the use cases will be industry / sector specific, or more general. I think there is a place for both types; however as a business analyst, I tend to write use cases for a specific stakeholder's requirements.

(11:31 AM) William Miller: is this list going out to the group?

(11:32 AM) Karen Guertler: Yes thanks.

(11:34 AM) Alicia Zuniga-Alvarado/AZA joined.

(11:36 AM) Dusty Jackson disconnected.

(11:36 AM) Tim Zimmerlin (Automation Technologies): Spatial Data ala Earth Observing System, Hubble Space Telescope, Google Maps, etc.

(11:36 AM) Karen Guertler: Excellent idea to incorporate use cases within the overall requirements work effort.

(11:36 AM) Yuri Demchenko (UvA) disconnected.

(11:37 AM) Tim Zimmerlin (Automation Technologies): Demographic Data ala Census, Metro Statistical Areas, Dept. Education, GDP, Employment, etc.

(11:38 AM) Bob Marcus38 joined.

(11:39 AM) Karen Guertler: Orit, I agree that Bob's input is very valuable. I do think that we have a few different perspectives as to what constitutes a use case, and what constitutes a potential technical implementation.

(11:40 AM) Tim Zimmerlin (Automation Technologies): Social Data ala Facebook, Google+, Linked In, Netflix Recommender, Amazon Customer Evaluations, etc.

(11:41 AM) Karen Guertler: The building blocks are very helpful; however they are a bit different from what I understand as requirements and use cases.

(11:42 AM) Karen Guertler: And, I agree with Bob wrt time constraints.

(11:42 AM) Yuri Demchenko (UvA) joined.

(11:42 AM) Tim Zimmerlin (Automation Technologies): Shopping Data ala Groupon, Expedia, Bizrate, etc.

(11:44 AM) Karen Guertler: I think we use different methodologies. I start with business requirements, including use cases. The technical implementations follow later in the methodology.

(11:44 AM) Yuri Demchenko (UvA): TMF published Big Data Analytics Model document where they described 16 use cases in a simple form as on epage table

(11:46 AM) William Miller: Goal of this group is to be technology agnostic

(11:47 AM) Bob Marcus20 joined.

(11:47 AM) Karen Guertler: Orit - yes - I really liked and understood Bob's proposal. Very clear. Just a different approach than I have used.

(11:49 AM) Bob Marcus20: Can we get a copy of TMF's Big Data work?

(11:50 AM) Karen Guertler: Bob, I agree; I think that the technical implementation is different from the use cases.

(11:51 AM) William Miller: we need to have better practice - best practice today has limitations that will not fulfill the vision for Big Data.

(11:51 AM) Karen Guertler: To me, the technical options derive from the business / organization use cases. I think we may be approaching this from a different POV.

(11:52 AM) Yuri Demchenko (UvA): It is the membership service but I will send you privately

(11:52 AM) William Miller: Comparision and constraint evalutation but most of all define the characteristic requirements and apply what best practice may be available or not avaialbe today

(11:54 AM) Karen Guertler: Bob's document is a good starting point.

(11:55 AM) Karen Guertler: Bob's comment just gets back to the scope of this work effort.

(11:55 AM) William Miller: an important requirement that has not been discussed is identification - the data needs to have a common way of idenification which can be translated into routability at the applicaitons layer

(11:57 AM) Yuri Demchenko (UvA): About identification, it must be one of the key reqs, in particular according to EU Open Data Initiative that defines PID for data and ORCID for researchers

(11:58 AM) Tim Zimmerlin (Automation Technologies): Geoffrey, please quickly give our group one use case to start.

(11:58 AM) Tim Zimmerlin (Automation Technologies): No, no, no: an actual documented use case in Google Docs.

(11:59 AM) William Miller: identifciaiton of the resource, device, type of data, all need a common means of identificaiton

(11:59 AM) Karen Guertler: I'm happy to provide a template / format...

(11:59 AM) William Miller: identification is also tied to security wich is handled by another subgroup

(11:59 AM) Karen Guertler: however, based on this discussion

(12:00 PM) Karen Guertler: I'm not sure that we have the same definition of 'use case'.

(12:00 PM) William Miller: define use case as read only or if the data is a bidirecitonal data source

(12:00 PM) Karen Guertler: I will post a template - hope it's helpful.

(12:01 PM) William Miller: we need to think about analytics and what makes Big Data Smart

(12:02 PM) Geoffrey Fox: i put algorithms needed as part of use case

(12:02 PM) Karen Guertler: I'd also like to confirm how this work effort relates to this: <http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.htm>

(12:03 PM) Karen Guertler: Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA)

(12:03 PM) Alicia Zuniga-Alvarado/AZA disconnected.

(12:03 PM) Karen Guertler: Wo, excellent example wrt medical data & PII.

(12:05 PM) Yuri Demchenko (UvA): We distinguish few different use domains for big data, at least; science, industry, business, living environment/cities, social mdeia and networks, healthcare

(12:06 PM) Tim Zimmerlin (Automation Technologies): Initially, use case bottlenecks are pivotal engineering and design information.

(12:07 PM) Geoffrey Fox:<http://grids.ucs.indiana.edu/ptliupages/publications/Where%20does%20all%20the%20data%20come%20from%20v7.pdf> has some use cases in descriptive fashion

(12:08 PM) Karen Guertler: Orit, I agree. Very broad\

(12:08 PM) Tim Zimmerlin (Automation Technologies): IMO, we can start each use case with data flows & storage resources.

(12:09 PM) Orit Levin (Microsoft): Karen, agreed.

(12:11 PM) Bob Marcus disconnected.

(12:12 PM) AliciaZuniga-Alvarado/AZA joined.

(12:14 PM) William Miller: sounds good