

NIST Big Data Public Working Group (NBD-PWG)/Subgroups Progress Report

Wo Chang, NIST
Robert Marcus, ET-Strategies
Chaitanya Baru, UC San Diego
<http://bigdatawg.nist.gov>

September, 4, 2013

AGENDA

- Review NBD-PWG Charter
- Review Subgroups Charter and Deliverables
- Review Overall Workplan
- Subgroup Current Activities Report
 - Definitions & Taxonomies Subgroup
 - Requirements Subgroup
 - Security & Privacy Subgroup
 - Reference Architecture Subgroup
 - Technology Roadmap Subgroup
- Expected Deliverables – 1st draft working drafts
- Face-to-face Meeting – September 30, at NIST
- Registration: https://www-s.nist.gov/CRS/conf_disclosure.cfm?&conf_id=6552
 - Deliverables Presentation & Discussion
 - Breakout Sessions by Subgroups
 - Announcement for Next Steps. Possible new work items:
 - White Paper: ReadMe First for NBD-PWG Documents
 - White Paper: Best Practices on Big Data Technologies and Solutions
 - *Others??*
- Q/A

SCOPE (Mo001)

The focus of the (NBD-PWG) is to form a community of interest from industry, academia, and government, with the goal of developing a consensus *definitions, taxonomies, secure reference architectures, and technology roadmap*. The aim is to create vendor-neutral, technology and infrastructure agnostic deliverables to enable big data stakeholders to pick-and-choose best analytics tools for their processing and visualization requirements on the most suitable computing platforms and clusters while allowing value-added from big data service providers and flow of data between the stakeholders in a cohesive and secure manner.

NBD-PWG

SCOPE AND DELIVERABLES

DELIVERABLES: Working Draft for

1. Big Data Definitions
2. Big Data Taxonomies
3. Big Data Requirements
4. Big Data Security and Privacy Requirements
5. Big Data Architectures Survey
6. Big Data Reference Architectures
7. Big Data Security and Privacy Reference Architectures
8. Big Data Technology Roadmap

LAUNCHED DATE:

June 26, 2013

TARGET DATE:

September 27, 2013

SUBGROUPS

AND THEIR SCOPES AND DELIVERABLES



Definitions and Taxonomies

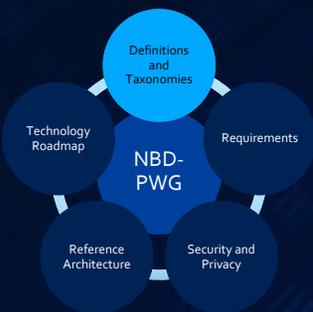
Nance Grady, SAIC
Natasha Balac, SDSC
Eugene Luster, R2AD

Scope (Moo18)

The focus is to gain a better understanding of the principles of Big Data. It is important to develop a consensus-based common language and vocabulary terms used in Big Data across stakeholders from industry, academia, and government. In addition, it is also critical to identify essential actors with roles and responsibility, and subdivide them into components and sub-components on how they interact/ relate with each other according to their similarities and differences.

Tasks

- For Definitions: Compile terms used from all stakeholders regarding the meaning of Big Data from various standard bodies, domain applications, and diversified operational environments.
- For Taxonomies: Identify key actors with their roles and responsibilities from all stakeholders, categorize them into components and subcomponents based on their similarities and differences
- Develop Big Data Definitions and taxonomies documents



Requirements and Use Cases

Geoffrey Fox, U. Indiana
Joe Paiva, VA
Tsegereda Beyene, Cisco



Scope (Moozo)

The focus is to form a community of interest from industry, academia, and government, with the goal of developing a consensus list of Big Data requirements across all stakeholders. This includes gathering and understanding various use cases from diversified application domains.

Tasks

- Gather input from all stakeholders regarding Big Data requirements.
- Analyze/prioritize a list of challenging general requirements that may delay or prevent adoption of Big Data deployment
- Develop a comprehensive list of Big Data requirements

Security and Privacy

Arnab Roy, CSA/Fujitsu
Nancy Landreville, U. MD
Akhil Manchanda, GE



Scope (Moo19)

The focus is to form a community of interest from industry, academia, and government, with the goal of developing a consensus secure reference architecture to handle security and privacy issues across all stakeholders. This includes gaining an understanding of what standards are available or under development, as well as identifies which key organizations are working on these standards.

Tasks

- Gather input from all stakeholders regarding security and privacy concerns in Big Data processing, storage, and services.
- Analyze/prioritize a list of challenging security and privacy requirements that may delay or prevent adoption of Big Data deployment
- Develop a Security and Privacy Reference Architecture that supplements the general Big Data Reference Architecture

Reference Architecture

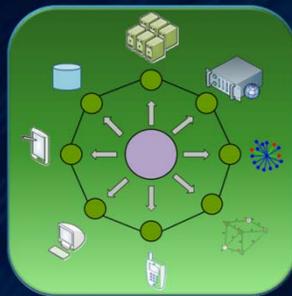
Orit Levin, Microsoft
James Ketner, AT&T
Don Krapohl, Augmented Intelligence

Scope (Moo21)

The focus is to form a community of interest from industry, academia, and government, with the goal of developing a consensus-based approach to orchestrate vendor-neutral, technology and infrastructure agnostic for analytics tools and computing environments. The goal is to enable Big Data stakeholders to pick-and-choose technology-agnostic analytics tools for processing and visualization in any computing platform and cluster while allowing value-added from Big Data service providers and the flow of the data between the stakeholders in a cohesive and secure manner.

Tasks

- Gather and study available Big Data architectures representing various stakeholders, different data types, use cases, and document the architectures using the Big Data taxonomies model based upon the identified actors with their roles and responsibilities.
- Ensure that the developed Big Data reference architecture and the Security and Privacy Reference Architecture correspond and complement each other.



Technology Roadmap

Carl Buffington, USDA/Vistronix
Dan McClary, Oracle
David Boyd, Data Tactic



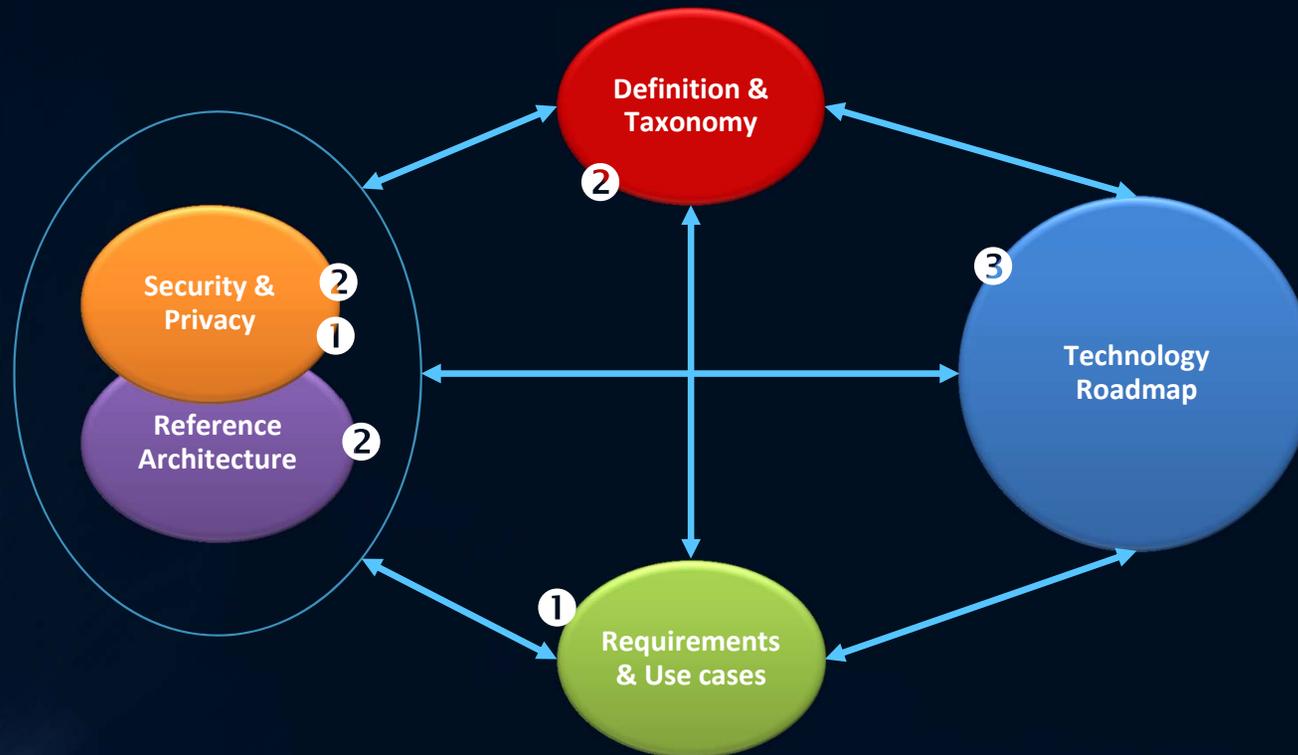
Scope (Moo22)

The focus is to form a community of interest from industry, academia, and government, with the goal of developing a consensus vision with recommendations on how Big Data should move forward by performing a good gap analysis through the materials gathered from all other NBD subgroups. This includes setting standardization and adoption priorities through an understanding of what standards are available or under development as part of the recommendations.

Tasks

- Gather input from NBD subgroups and study the taxonomies for the actors' roles and responsibility, use cases and requirements, and secure reference architecture.
- Gain understanding of what standards are available or under development for Big Data
- Perform a thorough gap analysis and document the findings
- Identify what possible barriers may delay or prevent adoption of Big Data
- Document vision and recommendations

Events Sequence & Information Flow Between Subgroups



Due to time constraints, activities will be carried out in parallel.

NIST Big Data Public Working Group and Subgroups Work Plan

Week	Def. & Tax.	Requirements	Sec. & Privacy	Ref. Arch	Tech. Roadmap
June 26	NBD-PWG (13:00PM – 15:00PM) Kick-off Meeting				
July 3	NBD-PWG (13:00PM – 15:00PM) Establish Subgroups with Co-Chairs, Subgroups Charter, Overall OWG direction				
July 8 - 12	Mondays 10:00AM – 12:00PM	Tuesdays 10:00AM – 12:00PM	Wednesdays 10:00AM – 12:00PM	Thursdays 10:00AM – 12:00PM	Fridays 10:00AM – 12:00PM
July 15 – 19	Definitions & Characteristics	Collect general use cases, identify requirements	Collect security and privacy use cases,	Analyze use cases from Reqs. & Sec. subgroups	Vision Characteristics & Def.
July 22 – 26	Tax.: Roles, activities, components & subcomp.	Categorize reqs., Identify missing reqs.	Identify requirements	Create conceptual model, identify actors,	Taxonomies Roles & Activities
July 24	NBD-WG (13:00PM – 15:00PM) Subgroups report: Sharing and brainstorming results				
July 29 – Aug. 2	↓	↓	↓	Identify usage scenarios, iden. Implement. Scenarios	Use cases & scenarios Ref. Architecture
Aug 5 – 9	↓	↓	↓	Create ref. architecture	Standards & Activities Gap Analysis
Aug 12 – 16	↓	↓	↓	↓	Standardization Priorities ??? Strategy of Adoption
Aug 19 – 23	↓	↓	↓	↓	Strategy of Implement. Resourcing
Aug. 21	NBD-WG (13:00PM – 15:00PM) Subgroups report: Present and Discuss Working Draft Outline				
Aug. 26 - 30	↓	↓	↓	↓	Recommendations
Sept. 2 – 6	↓	↓	↓	↓	↓
Sept. 4	NBD-WG (13:00PM – 15:00PM) Subgroups report: Present and Discuss Rough Draft				
Sept 9 – 13	↓	↓	↓	↓	↓
Sep 16 - 20	↓	↓	↓	↓	↓
Sep 23 – 27	↓	↓	↓	↓	↓
Sep 25	NBD-WG (13:00PM – 15:00PM) Subgroups report: Present and Discuss Final Draft				
Sep 30	Big Data Workshop, NIST - Deliverables Presentation & Discussion - Breakout Sessions by Subgroups - Announcement for Next Steps				

NIST Big Data

NIST Big Data Program

Welcome to NIST Big Data Public Working Group (NBD-PWG)!

Search



Home

- [NBD-WG/Subgroups](#)
- [Charter](#)
- [Co-Chairs](#)
- [Guidelines](#)
- [Work Plan](#)
- [Virtual Meeting](#)

Tasks: To Develop

- [Definitions](#)
- [Taxonomies](#)
- [Requirements](#)
- [Ref. Architectures](#)
- [Tech. Roadmap](#)

Documents

- [Input Listing](#)
- [Output Listing](#)
- [Upload Document](#)

Registration

- [New User](#)
- [Update Profile](#)

Points of Contact

Chris Greer
NIST / ITL
Associate Director,
Program Implementation

Wo Chang

NIST / ITL
Digital Data Advisor

Hot News:

- [NIST's 1st Big Data Working Group Meeting](#) [- Rescheduled - Via Teleconferencing, June 26, 2013, 13:00 - 15:00 EDT]
- [NIST's 1st Big Data Working Group Meeting](#) [Via Teleconferencing, June 19, 2013, 13:00 - 14:00 EDT]

Input Document Listing - Entire

Documents Received: 156

Please review and prepare to discuss the submitted documents before each meeting.

[Entire](#) | [All](#) | [DefTax](#) | [Requirements](#) | [SecPrivacy](#) | [RefArch](#) | [TechRoadmap](#) | [UseCases](#) | [Gap](#)

No.	Date	Subgroup	Title	Revision
M0156	2013-08-22	All	NBD-PWG/Subgroups Joint Meeting Minutes for Aug. 21, 2013	v1
M0155	2013-08-22	UseCases	EISCAT 3D incoherent scatter radar system Use Case	v1
M0154	2013-08-22	RefArch	Meeting Agenda 8/22/2013	v1
M0153	2013-08-22	RefArch	Meeting Summary from Aug 15, 2013	v1
M0152	2013-08-21	Requirements	Requirements Working Draft	v1
M0151	2013-08-21	RefArch	White Paper "Big Data Reference Architectures"	v1
M0150	2013-08-21	All	NBD-PWG/Subgroups Joint Meeting PPT for Aug. 21, 2013	v1
M0149	2013-08-21	All	Guidance on Using Big Data WG Deliverables	v1
M0148	2013-08-21	UseCases	National Archives and Records Administration Accession NARA, Search, Retrieve, Preservation (Digital Archives) Use Case	v1
M0147	2013-08-21	UseCases	Census 2010 and 2000 - Title 13 Big Data (Digital Archives) Use Case	v1
M0146	2013-08-20	TechRoadmap	Align_BigData_Roles_v1	v1
M0145	2013-08-20	TechRoadmap	Alignment_RA_and_Roadmap_v1	v1
M0144	2013-08-20	All	Agenda for NBD-PWG/Subgroups Joint Meeting for Aug. 21, 2013	v1
M0143	2013-08-20	Requirements	NBD-Requirements WG Meeting Minutes August 20 2013	v1
M0142	2013-08-20	DefTax	Subgroup 1: Def/Tax draft	v1
M0141	2013-08-20	UseCases	Use case: Biodiversity (UvA)	v1
M0140	2013-08-19	UseCases	Individualized Diabetes Management Use Case	v1
M0139	2013-08-19	Requirements	NBD-Requirements WG Agenda August 20, 2013	v1 , v2
M0138	2013-08-19	Requirements	Draft Requirements Report	v1

Upcoming Events

- [NIST Big Data Workshop, NIST, September 30, 2013](#)

Useful References

- [Frontiers in Massive Data Analysis, July, 2013](#)
- [Obama Administration Unveils "Big Data" Initiative; Announces \\$200 Million in New R&D Investments, March 29, 2012](#)
- [Big Data: The next frontier for innovation, competition, and productivity, June 2011](#)

Progress: Definitions and Taxonomies Subgroup (Mo024, Mo142)

Key documents: Mo024 – ongoing discussion, Mo142 – working draft

- Big Data Definitions, v1 (Developed from Jan. 15 - 17, 2013 NIST Cloud/BigData Workshop)

Big Data refers to digital data volume, velocity and/or variety that:

- *enable novel approaches to frontier questions previously inaccessible or impractical using current or conventional methods; and/or*
- *exceed the storage capacity or analysis capability of current or conventional methods and systems.*
- *differentiates by storing and analyzing population data and not sample sizes.*

Data Scientists (New) is a practitioner who has sufficient knowledge of the overlapping regimes of expertise in domain knowledge, analytical skills and programming expertise to manage the analytics process through each stage in the big data lifecycle. They handle value, veracity (quality), etc.

Progress: Definitions and Taxonomies Subgroup (Mo024, Mo142)

Key documents: Mo024 – ongoing discussion, Mo142 – working draft

■ Big Data Taxonomies (Mo202)



Progress: Requirements and Use Case Subgroup

Key documents: Mo105 – use cases, Mo125, requirements, Mo152 – working draft

- Use Case Template

1. *Goals, Description*
2. *Data Characteristics, Data Types*
3. *Data Analytics*
4. *Current Solutions*
5. *Security & Privacy*
6. *Lifecycle Management and Data Quality*
7. *System Management and Other issues*

- 45 Use Cases Received

<http://bigdatawg.nist.gov/usecases.php>

1. *Government Operation (2)*
2. *Commercial (8)*
3. *Healthcare and Life Sciences (10)*
4. *Deep Learning and Social Media (6)*
5. *Ecosystem and Research (4)*
6. *Astronomy and Physics (5)*
7. *Earth, Environmental and Polar Science (10)*

Use Case Title		
Vertical (area)		
Author/Company/Email		
Actors/Stakeholders and their roles and responsibilities		
Goals		
Use Case Description		
Current Solutions	Compute(System)	
	Storage	
	Networking	
	Software	
Big Data Characteristics	Data Source (distributed/centralized)	
	Volume (size)	
	Velocity (e.g. real time)	
	Variety (multiple datasets, mashup)	
	Variability (rate of change)	
Big Data Science (collection, curation, analysis, action)	Veracity (Robustness Issues, semantics)	
	Visualization	
	Data Quality (syntax)	
	Data Types	
Big Data Specific Challenges (Gaps)		
Big Data Specific Challenges in Mobility Security & Privacy Requirements		
Highlight issues for generalizing this use case (e.g. for ref. architecture)		
More Information (URLs)		
Note: <additional comments>		
Note: No proprietary or confidential information should be included		

Progress: Security and Privacy Subgroup

Key documents: Google Doc – ongoing discussion, Mo110 – requirements working draft, Moxxxx – architecture & taxonomies

■ Requirements Scope

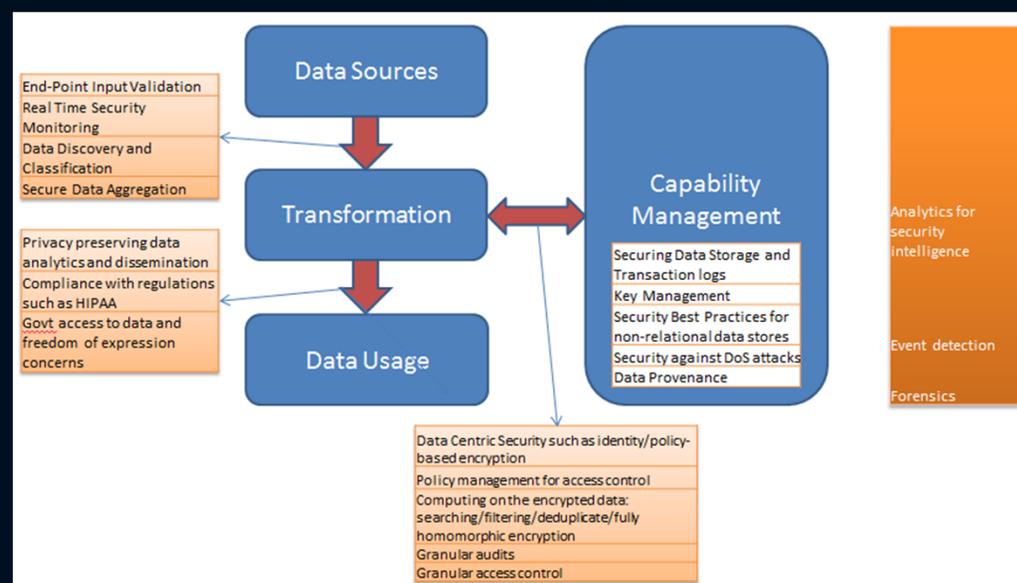
1. *Infrastructure Security*
2. *Data Privacy*
3. *Data Management*
4. *Integrity and Reactive Security*

■ Requirements – Use Cases Studied

1. *Retail (consumer)*
2. *Healthcare*
3. *Media (social media and communications)*
4. *Government (military, justice systems, etc.)*
5. *Marketing*

■ Architecture & Taxonomies

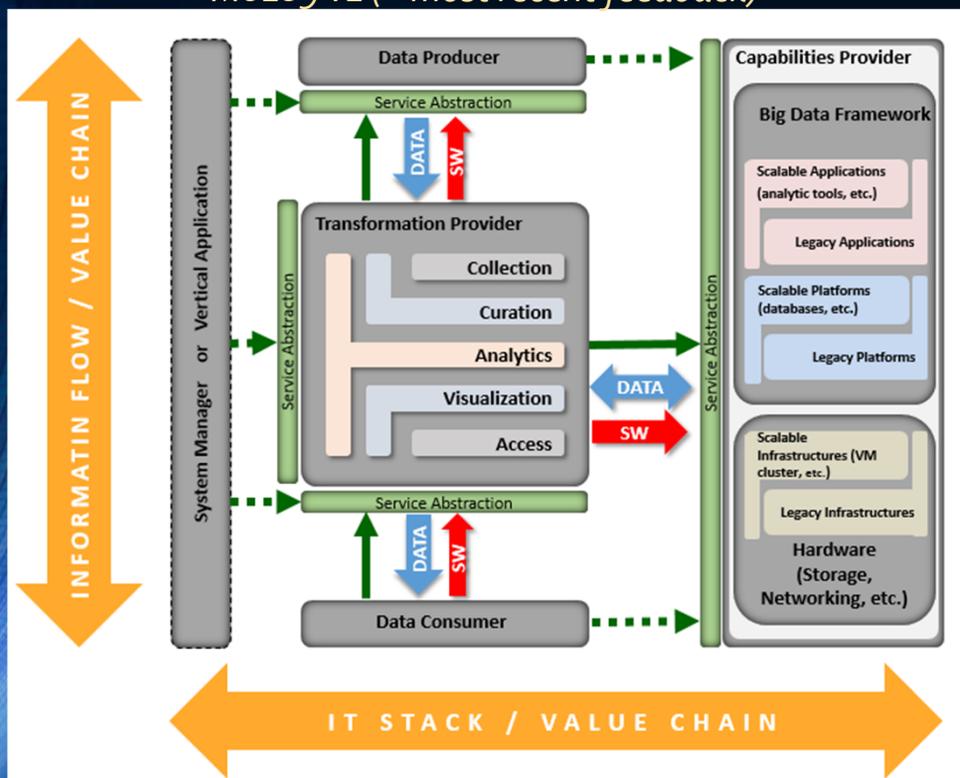
1. *Privacy*
2. *Provenance*
3. *System Health*



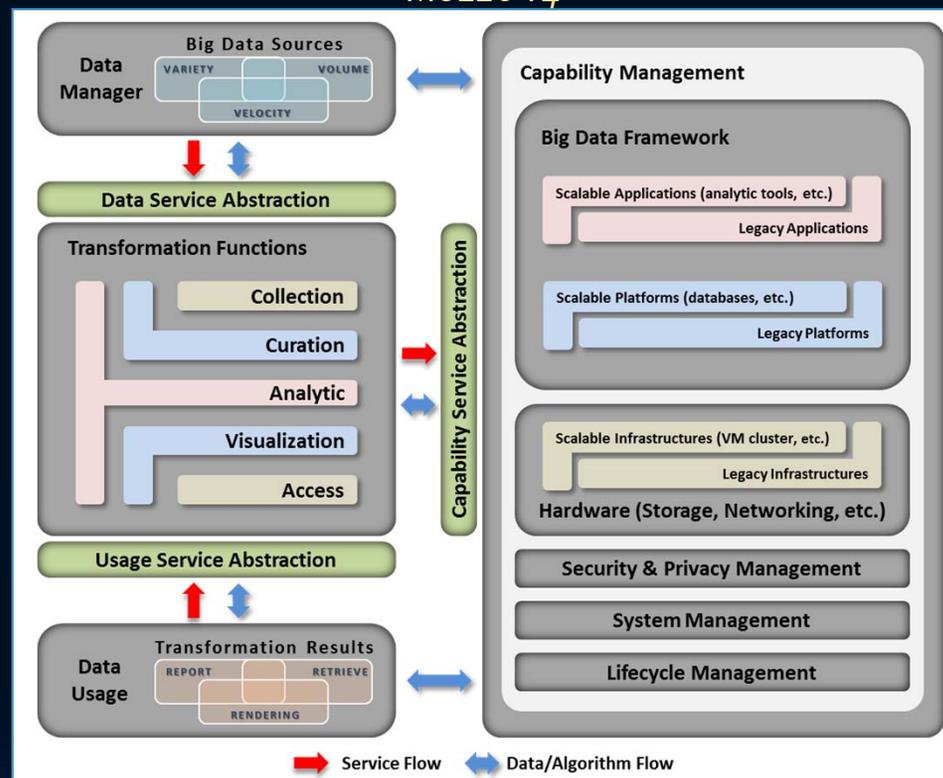
Progress: Reference Architecture Subgroup

Key documents: Mo100 – ongoing discussion, Mo151 – white paper, Mo123 working draft

Mo189 v1 (+ most recent feedback)



Mo126 v4



Progress: Technology Roadmap Subgroup

Key documents: Moo87 – working draft

- Inputs from other subgroups
 1. *Definitions and Taxonomies*
 2. *Requirements and Use Cases*
 3. *Security and Privacy*
 4. *Reference Architecture*
- Potential Standards Group with Big Data related activities (Moo35)
- Capabilities and Technology Readiness
- Big Data Decision Framework
- Big Data Mapping and Gap Analysis
- Big Data Strategies
 1. *Adoption*
 2. *Implementation*
 3. *Resourcing*

Subgroups Working Draft Outline

- Definitions & Taxonomies (Mo142)
- Requirements and Use Cases (Mo152)
- Security & Privacy (Mo110 – requirements; Moxxx – architecture)
- Reference Architecture (Mo151 – white paper; Mo123 – working draft)
- Technology Roadmap (Moo87)

<i>Progress...</i>	July (brainstorm)	Aug (outline)	Sep (write-up)
Definitions and Taxonomies	<i>(1) For Definitions: Compile terms used from all stakeholders regarding the meaning of Big Data from various environments. (2) For Taxonomies: Identify key actors with their roles and responsibilities from all stakeholders, categorize them into components and subcomponents based on their similarities and differences</i>		
Requirements and Use Cases	<i>(1) Gather input from all stakeholders regarding Big Data requirements. (2) Analyze/prioritize a list of challenging general requirements that may delay or prevent adoption of Big Data deployment. (44 Use cases received; ~26 general requirements in 7 categories)</i>		
Security and Privacy	<i>(1) Gather input from all stakeholders regarding security and privacy concerns in Big Data processing, storage, and services. (2) Analyze/prioritize a list of challenging security and privacy requirements that may delay or prevent adoption of Big Data deployment .</i>		
Reference Architecture	<i>(1) Gather and study available Big Data architectures representing various stakeholders, different data types, use cases, and document the architectures using the Big Data taxonomies model based upon the identified actors with their roles and responsibilities. (2) Ensure that the developed Big Data reference architecture and the Security and Privacy Reference Architecture correspond and complement each other.</i>		
Technology Roadmap	<i>(1) Gather input from other subgroups and study the taxonomies for the actors' roles and responsibility, use cases and requirements, and secure reference architecture. (2) Gain understanding of what standards are available or under development for Big Data. (3) Perform a thorough gap analysis and document the findings. (4) Identify what possible barriers may delay or prevent adoption of Big Data.</i>		

NIST BIG DATA PWG NEXT STEPS:

- Possible New Work Items:
 - *White Paper: ReadMe First for NBD-PWG Documents*
 - *White Paper: Best Practices on Big Data Technologies and Solutions*
 - *Others??*

BIG DATA WORKSHOP (FACE-TO-FACE MEETING):

Date: September 30, 2013

Location: NIST

Registration: https://www-s.nist.gov/CRS/conf_disclosure.cfm?&conf_id=6552