

# **NIST Big Data Public Working Group**

## **Overview of NIST Big Data Interoperability Framework Volume 3**

### **Prof. Geoffrey Fox**

Distinguished Professor of  
Computer Science and Informatics  
Indiana University

### **Dr. Piyush Mehrotra**

Chief, NASA Advanced  
Supercomputing Division  
NASA

**NIST Campus**

**Gaithersburg, Maryland**

**June 1, 2017**

# Presentation Overview

- Volume Presentation Outline
- Volume 1, Definitions (Nancy Grady, SAIC)
- Volume 2, BD Taxonomies (Nancy Grady, SAIC)
- **Volume 3, Use Cases and General Requirements (Geoffrey Fox, Indiana University)**
- Volume 6, Reference Architecture (David Boyd, InCadence Corp.)
- Volume 4, Security and Privacy (Arnab Roy, Fujitsu; Mark Underwood, AVP, Strategic Initiatives, Controls and Countermeasures)
- Volume 8, Reference Architecture Interface (Gregor von Laszewski, Indiana University)
- Reference Architecture Software Implementation Environment and Demonstration (Gregor von Laszewski, Indiana University)
- Volume 7, Standards Roadmap (Russell Reinsch, Center for Government Interoperability)
- Volume 9, Adoption and Modernization (Russell Reinsch, Center for Government Interoperability)

# NBDIF Volume Overview

**Vol. 1 BD Definitions**  
Defines common language

**Vol. 2 BD Taxonomies**  
Hierarchy of NBDRA components

**Vol. 3 Use Cases & Vol. 5 Arch Survey**  
Info gathered; requirements extracted

**Vol. 6 NBDRA**  
Developed NBDRA

**Vol. 4 S&P**  
Interwoven topics of S&P examined

**Vol. 7 Standards Roadmap**  
Examine standards wrt NBDRA

**Vol. 8 NBDRA Interfaces**  
Implementation of NBDRA

**Vol. 9 Adoption & Modernization**



# Volume Presentation Outline

---

- For each volume
  - Scope of the volume
  - Brief recap of version 1
  - Highlights of version 2 accomplishments
  - Summary of version 2 areas needing contributions
  - Topics that could be considered for version 3

## Volume 3, Use Cases and General Requirements

# Document Scope

---

- **Version 1 collected 51 big data use cases with a 26 feature template and used this to extract requirements to feed into NIST Big Data Reference Architecture**
- **The version 2 template merges version 1 General and Security & Privacy use case analysis**
- **The discussion of this at first NIST Big Data meeting identified need for patterns which were proposed during version 2 work; version 2 template incorporates new questions to help identify patterns.**

## Volume 3, Use Cases and General Requirements

# Version 1 Overview

---

- Gathered and evaluated **51** use cases from nine application domains.
- Gathered input regarding Big Data requirements
- Analyzed and prioritized a list of challenging use case specific requirements that may delay or prevent adoption of Big Data deployment
- Developed a comprehensive list of generalized Big Data requirements
- Developed a set of features that characterized applications
  - Used to compare different Big Data problems
- Collaborated with the NBD-PWG Reference Architecture Subgroup to provide input for the NBDRA

# 51 Detailed Use Cases: Version 1 Contributed July-September 2013

<http://bigdatawg.nist.gov/usecases.php>, 26 Features for each use case

- **Government Operation(4):** National Archives and Records Administration, Census Bureau
- **Commercial(8):** Finance in Cloud, Cloud Backup, Mendeley (Citations), Netflix, Web Search, Digital Materials, Cargo shipping (as in UPS)
- **Defense(3):** Sensors, Image surveillance, Situation Assessment
- **Healthcare and Life Sciences(10):** Medical records, Graph and Probabilistic analysis, Pathology, Bioimaging, Genomics, Epidemiology, People Activity models, Biodiversity
- **Deep Learning and Social Media(6):** Driving Car, Geolocate images/cameras, Twitter, Crowd Sourcing, Network Science, NIST benchmark datasets
- **The Ecosystem for Research(4):** Metadata, Collaboration, Translation, Light source data
- **Astronomy and Physics(5):** Sky Surveys including comparison to simulation, Large Hadron Collider at CERN, Belle II Accelerator in Japan
- **Earth, Environmental and Polar Science(10):** Radar Scattering in Atmosphere, Earthquake, Ocean, Earth Observation, Ice sheet Radar scattering, Earth radar mapping, Climate simulation datasets, Atmospheric turbulence identification, Subsurface Biogeochemistry (microbes to watersheds), AmeriFlux and FLUXNET gas sensors
- **Energy(1):** Smart grid





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	
	Data Analytics	
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>		

# Version 1 Use Case Template

- Note agreed in this form August 11 2013
- Some clarification on Veracity v. Data Quality added
- Request for picture and summary done by hand for version 1 but included in version 2 template.
- Early version 1 use cases did a detailed breakup of workflow into multiple stages which we want to restore but do not have agreed format yet



## Volume 3, Use Cases and General Requirements

# Version 2 Accomplishments I

---

- **Developed the Use Case Template version 2 to describe additional use cases with more specific information**
  - Extra information developed from looking at 51 version 1 use cases and seeing where was missing or vague
  - Made available as a PDF or Google Form – Version 1 a Word document
- **Version 1 use case combined with Security and Privacy (S&P) analysis to generate very detailed information on this**
  - Can also be used to generate only S&P information or only general information such as analysis environment and data characteristics
- **Continued study of characteristics of version 1 use cases**
  - compare Big Data with High Performance Computing workloads
- **Will evaluate additional use cases as they are submitted**

## Volume 3, Use Cases and General Requirements Version 2 Accomplishments II

- **Socialized Version 1 use cases with the developing ISO Big Data community, which reportedly found them representative and useful**
- **The new template was developed to:**
  - **Better understand processing stages and workflow**
  - **Gather additional use case detail**
  - **Collect use cases with sensitive privacy & other security concerns (sidebar)**
  - **Two versions of the new template were deployed**
- **Collecting detail-rich use cases is challenging**
- **Added key use cases from US Census and health care (consent flow and “Break Glass”)**

- Roles
- PII
- Covenants, Liability
- Ownership, Identity, Distribution
- Risk Mitigation
- Provenance
- Data Life Cycle
- Audit and Traceability
- Application Provider Security
- Framework Provider Security
- System Health
- Permitted Use Cases

## Volume 3, Use Cases and General Requirements Version 2 Opportunities for Contribution

- More use cases. (Roll up sleeves; budget an hour.)
- Soliciting greater application domain diversity:
  - Smart cars (Smart X)
  - Large scale utility IoT
  - Geolocation applications involving people
  - Energy from discovery to generation
  - Scientific studies involving human subjects at large scale
  - Highly distributed use cases bridging multiple enterprises



**Choose a domain and collect/analyze a set of related use-cases**

**Develop technology requirements for applications in domain**  
**Feed lessons into version 3 of template**

## Volume 3, Use Cases and General Requirements

# Possible Version 3 Topics

- Identify gaps in use cases
- Develop plausible, semi-fictionalized use cases from industry reports, white papers, academic project reports
- Identify important parameters for estimating systems
- Microservice use cases
- Use cases mapped to work in Vol 8
- Container-oriented use cases
- Forensic and provenance-centric use cases
- Review fitness of the BDRA to use cases

