

NIST Big Data Public Working Group

Overview of NIST Big Data Interoperability Framework Volume 7

Russell Reinsch

Analyst

Center for Government Interoperability

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 7, Standards Roadmap

Document Scope

- Provide
 - an agnostic resource for organizations to review the current state of big data standards
 - a generic outlook on potential impacts of standards development in various areas
- Conduct analysis of where closure of standards gaps may benefit economic and social development



Volume 7, Standards Roadmap Version 1 Overview

- Reviewed work from the other six subgroups
- Cataloged a collection of existing Big Data related standards
- Mapped collected standards to NBDRA components
- Identified 16 significant gaps in existing standards, highlighting areas where standards are expected to have significant impact in the future
- Discussed a pathway to address standards gaps

Volume 7, Standards Roadmap

Version 2 Accomplishments

- Clarified alternative perspectives for viewing standards (functional vs. organizational; product vs. non-product)
- Applied a mapping technique for pairing standards to requirements which were defined in Volume 3
- Applied a mapping technique for pairing standards to selected use cases defined in Volume 3

Mapping Standards to Specific Requirements from Vol 3

Table 2: Data source requirements-to-standards matrix:

Requirement	Requirement Description	Standard Description	Standard
DSR-1	Fast search		
DSR-2			
DSR-3	Visual layout of results for presentation.	Suggested charts and tables for various purposes.	IBCS notation; related: ACRL
DSR-4	Browser access		<u>WebRTC</u>
DSR-5	Layer standard		ISO 13606
DSR-6			

Mapping Standards to Use Cases from Volume 3

Table 3: An excerpt of the master use case-to-standards matrix:

Use Case Number and Type	Use Case Description	Standard Description	Standard
2: Government	Information retrieval / records search in US Census DB		
6: Commercial	Research DB document recommender, impact forecast		
8: Commercial	Web search	See table 4, M0165 breakout supplement on next page. Query [no SY]:	<u>Xpath</u> , <u>xquery</u> , <u>full-text</u> , <u>elixir</u> , <u>xirql</u> , <u>xml</u> .
15: Defense	Intelligence data processing		
34: Research	Graph DB search		

Sections of Use Cases Applied to Mapping Breakouts

NIST Big Data (NBDWG) Use Case Requirements Template

Use Case Title		
Vertical (area)		
Author/Company		
Actors/Stakeholders and their roles and responsibilities		
Goals		
Use Case Description		
Current Solutions	Compute(System)	
	Storage	
	Analytics(Software)	
Big Data Characteristics	Volume (size)	
	Velocity	
	Variety	
	Veracity (Robustness Issues)	
	Visualization	
	Data Quality	
Big Data Specific Challenges (Gaps)		
Security & Privacy Requirements		
More Information (URLs)		
Note: <additional comments>		

Mapping Existing Standards to Use Case # 8

Table 4: M0165 Breakout Supplement

Current solution	Compute system	Large cloud	
Current solution	Storage	Inverted index	
Current solution	Networking	External most important	SRU, SRW, [CQL], z39.50; OAI PMH; <u>Spargl</u> , REST, <u>Href</u> ;
Current solution	Software		Spark
Data science	Veracity	Main hubs, authorities	
Data science	Visualization	Page layout is critical. Technical elements inside a site affect content delivery.	
Data science	Data quality		<u>SRank</u>
Data science	Data types		
Data science	Data analytics	Crawl, preprocess, index, rank, cluster, recommend. Crawling / collection: connection elements including mentions from other sites	Sitemap.xml, responsive design,
Gaps		Links to user profiles, social data	Schema.org

Volume 7, Standards Roadmap

Version 2 Opportunities for Contribution

- **Expand discussion on standards gaps (from the list of 16) and the impact standards could have on that area (Section 4.4)**
- **Contribute text for discussion of standards for integration (also Section 4.4)**
- **Contribute to the catalog of existing data standards (Section 5.2)**
- **Help complete mapping standards to requirements and standards to use cases (Sections 2.3.1 & 2.3.2)**



Volume 7, Standards Roadmap

Possible Version 3 Directions

- Investigate a number of additional areas including:
 - Internet of Things [IoT] impact
 - Network connectivity
 - Complex Event [CEP] and real-time processing
 - Data Marketplaces
- Create SV-1 / CV-6 frameworks for viewing system and component incompatibilities and interoperability.
- Conform to and integrate with NIST BDPWG User Guide
- Discover guiding principles, practical strategies and suggested sequences of actions for future efforts

CV-6 for Mapping Activities and Capabilities to Reference Architecture

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	Operational Activities															
2	Capability to Operational Activities Mapping	Collection	Preparation	Analytics	Visualization	Access	Resource Mgmt	Messaging	Batch Processing	Interactive Processing	Stream Processing	Index	Read	Update	authorization	
3	orchestrate dependencies															
4	authenticate kerberos															
5	automatic refresh of dash objects															
6	re sort in dashboard [root cause] drill from report to underlying source															
7	search on metadata in report repository															
8	graphically present sentiments						Google chart									
9	Capabilities															
10		classify			sentiment analysis											
11		classify polarity									iterate: max entropy					
12		classify polarity			naïve bayes: categorize text											
13		classify polarity			baseline against											
14		train classifier			evaluate classifier											
15		calculate pt to nt ratio			set threshold											
16		pre process		lingpipe: label					distribute data							
17		pre process		regular expression												
18		pre process		split dataset												
19	pre process		remove urls, handles													
20	pre process		coding required													
21	channel connector	retrieve tweets					twitter api									twitter api
22	channel connector						python streaming api									
23																
24																
25																
26																