**DRAFT User Guide**

The following document is a draft of a user guide to the nine NIST Big Data Interoperability Framework (NBDIF) volumes. The objective of this document is to provide a very high level overview of the contents of the nine volumes, relation between the volumes, location (within the volumes) of some topics of interest to the Big Data community, and related NIST resources. The proposed focus of each section is highlighted in blue. The blue text serves to guide development of the document and will be removed later. Example text (i.e., text similar to what will be included in the document) is not highlighted.

# Volume Overview

## OVERALL GOAL OF THE NBDIF

This section will describe (at a very high level) the purpose for creating the compendium of NBDIF documents (e.g., Why were the documents created? What do the documents contain?)

## Suggested Reading Pathways

This section will suggest a reading pathway (i.e., a suggested order in which to read the volumes) for a few user groups. The user groups are very high-level, broad groups and are not intended to address every scenario of user.

One goal of the User Guide is to assist the reader in finding information of interest within the volumes of the NBDIF. Readers from different disciplines may prefer to take different navigation paths on their journey through the volumes. The suggested reading pathways below are only intended as a guide and do not suggest that these are the only order to read the NBDIF volumes.

Analyst (e.g., Security and Privacy Analysts, Vertical Application End-Users)

Suggested reading pathway narrative

Business User (e.g., Project Managers, Senior Executives)

Suggested reading pathway narrative: Vol 3 🡪 Vol n 🡪 Vol n2.

Engineer (e.g., Chief Data Officers, Database Managers, Developers, IT Staff, System Architects

Suggested reading pathway narrative

Scientist (e.g., Data Scientists, Vertical Application End-Users)

Suggested reading pathway narrative

Standards Developer

Suggested reading pathway narrative

## Volume Abstracts

This section describes at a very high level (i.e., 2-4 sentences) the focus of each volume.

# NBDIF Index and Thesaurus

This section will list major topics and subtopics covered within the volumes and point users to the location of significant discussion of the subtopic within the volumes.

## Index

| Topic  Subtopic | | Vol 1 | Vol 2 | Vol 3 | Vol 4 | Vol 5 | Vol 6 | Vol 7 | Vol 8 | Vol 9 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ANALYTICS | |  |  |  |  |  |  |  |  |  |
|  | application |  |  | pg 44, 50, 289, rerun |  |  |  |  |  |  |
|  | as a central concept in the reference architecture |  |  |  |  |  |  | pg 21, 22 |  |  |
|  | as a general requirement characteristic |  |  |  |  |  |  | pg 16 |  |  |
|  | as a key component of use case documentation |  |  |  |  |  |  | pg 30, 31, 32 |  |  |
|  | as it relates to the gaps |  |  |  |  |  |  | pg 35, 40, 43, 49, 50 |  |  |
|  | as part of a workflow |  |  | pg 268, 274 |  |  |  |  |  |  |
|  | benefit of |  |  | pg 327 |  |  |  |  |  |  |
|  | current approaches |  |  | pg 9, 14, 15, 17, 327 |  |  |  |  |  |  |
|  | definition |  |  | pg 269 |  |  |  |  |  |  |
|  | facets of |  |  | pg 272 |  |  |  |  |  |  |
|  | future needs |  |  | pg 8, 23, 25, 26, 50, 327 |  |  |  |  |  |  |
|  | in relation to security and privacy |  |  |  |  |  |  | pg 19, 20 |  |  |
|  | pertaining to data cleaning and quality |  |  |  |  |  |  | pg 47, 49 |  |  |
|  | relationship with transformation |  |  | pg 51 |  |  |  |  |  |  |
|  | types of |  |  | pg 329 |  |  |  |  |  |  |
|  | UIMA |  |  |  |  |  |  | pg 64 |  |  |
| CLEANING | |  |  |  |  |  |  |  |  |  |
|  | during migration, |  |  |  |  |  |  | pg 42 |  |  |
|  | as it relates to quality |  |  |  |  |  |  | pg 47 |  |  |
|  | time sink |  |  |  |  |  |  |  |  | pg 39, 41 |
|  | pervasive project issue |  |  |  |  |  |  |  |  | pg 25 |
|  | position in a ML workflow |  |  |  |  |  |  |  |  | pg 28 |
|  | usecase 8, M0165 (recheck) |  |  | usecase 8, M0165 |  |  |  |  |  |  |
|  | usecase 2-1 (recheck) |  |  | usecase 2-1 |  |  |  |  |  |  |
|  | usecase 41 (recheck) |  |  | usecase 41 |  |  |  |  |  |  |
|  | usecase 51 (recheck) |  |  | usecase 51 |  |  |  |  |  |  |
|  | relationship to schema on read | sec 2 |  |  |  |  |  |  |  |  |
|  | relationship to preparation | sec 5.3 |  |  |  |  |  |  |  |  |
|  | differing levels of importance | sec 5 |  |  |  |  |  |  |  |  |
| [DATA] PREP | |  |  | zero |  |  |  |  |  |  |
|  | as part of the data life cycle |  |  |  |  |  |  | pg 15 |  |  |
|  | in the context of 'translation' |  |  |  |  |  |  | pg 17 |  |  |
|  | time sink |  |  |  |  |  |  | pg 48 |  |  |
|  | in context of NBDRA Activity |  |  |  |  |  |  | pg 77 |  |  |
|  | in the context of FAIR |  |  |  |  |  |  |  |  | pg 0 and sec 6.3.5 |
|  | in the context of governance |  |  |  |  |  |  |  |  | sec 6.3.4 |
|  | time sink |  |  |  |  |  |  |  |  | sec 6.2.1.6 |
|  | related to implementation |  |  |  |  |  |  |  |  | sec 6.3.5 |
|  | importance to checklist design |  |  |  |  |  |  |  |  | sec 6.3.6.2 |
|  | level of priority, checklist ID 2d?? |  |  |  |  |  |  |  |  | level of priority, checklist ID 2d?? |
|  | fundamental part of the analytics lifecycle | section 2, 5.3 |  |  |  |  |  |  |  |  |
|  | in relation to tranformation, cleaning, integration, and schema on read | sec 2 |  |  |  |  |  |  |  |  |
|  | changes to the order of steps in processing | sec 5.3.1 |  |  |  |  |  |  |  |  |
| GRAPHS | |  |  |  |  |  |  |  |  |  |
|  | knowledge graphs and RDF |  |  |  |  |  | graph DB, pg 43 | pg 37, 39 |  |  |
|  | data structure |  |  |  |  |  |  | pg 41 |  |  |
|  | usecase 34 (recheck) |  |  | usecase 34, 6 hits |  |  |  |  |  |  |
|  | usecase 21 (recheck) |  |  | usecase 21, 3 hits |  |  |  |  |  |  |
|  | usecase 32 (recheck) |  |  | usecase 32, concept graph |  |  |  |  |  |  |
|  | usecase 51 (recheck) |  |  | usecase 51, pg 191 |  |  |  |  |  |  |
|  | support for graph data and processing |  |  | sec 3.2 |  |  |  |  |  |  |
|  | a class12n dimension for mapping the 51 usecases |  |  | appendix E |  |  |  |  |  |  |
|  | as a vis12n technique, 1.19?? |  |  | as a vis12n technique, 1.19?? |  |  |  |  |  |  |
|  | usecase 2-3 (recheck) |  |  | usecase 2-3 |  |  |  |  |  |  |
|  | graph DB | sec 4.2.2 |  |  |  |  |  |  |  |  |
|  | visualization challenges | sec 3.3.5 |  |  |  |  |  |  |  |  |
| LANGUAGE | |  |  |  |  |  |  |  |  |  |
|  | related to English and terminology |  |  |  |  |  |  | pg 14 |  |  |
|  | gaps affecting interoperability |  |  |  |  |  |  | pg 35, 52 |  |  |
|  | related to query |  |  |  |  |  |  | pg 39, 40, 41 |  |  |
|  | when accepted by a service |  |  |  |  |  |  | pg 28 |  |  |
|  | in geo spatial mapping |  |  |  |  |  |  | pg 41 |  |  |
|  | related acronyms |  |  |  |  |  |  | pg 53, 54, 55, 56 |  |  |
|  | related standards |  |  |  |  |  |  | appendix B |  |  |
|  | NLP |  |  | usecase 15, 22, 31, 2-3 |  |  |  |  |  |  |
|  | XBRL |  |  | usecase 5 |  |  |  |  |  |  |
|  | KML |  |  | usecase 13 |  |  |  |  |  |  |
|  | XML |  |  | usecase 29 |  |  |  |  |  |  |
|  | stemming |  |  | usecase 34 |  |  |  |  |  |  |
|  | ENVRI [is a] reference model |  |  | **usecase 42** |  |  |  |  |  |  |
|  | semistructured data, XML, JSON | sec 3.2.5, sec 3.31 |  |  |  |  |  |  |  |  |
|  | issue with the term nosql | sec 4.2.2 |  |  |  |  |  |  |  |  |
|  | new programming languages [sql layers?] | sec 4.4.6 |  |  |  |  |  |  |  |  |
|  | related acronyms [3] | appendix A |  |  |  |  |  |  |  |  |
| MASHUP | |  |  |  |  |  |  |  |  |  |
|  | metadata issues in mashup |  |  |  |  |  |  |  |  | pg 38 |
|  | in commercial usecases |  |  | pg 90, 91, 118, 210, 240 |  |  |  |  |  |  |
|  | benefit |  |  | pg 8 |  |  |  |  |  |  |
|  | relationship with Variety |  |  | pg 268 |  |  |  |  |  |  |
|  | relationship to linked data | sec 5.4.6 |  |  |  |  |  |  |  |  |
|  | [neg] effect on security and privacy | sec 5.5.2 |  |  |  |  |  |  |  |  |
| METADATA | |  |  |  |  |  |  |  |  |  |
|  | synonymous with descriptive data |  |  |  |  |  |  | pg 14 |  |  |
|  | its role in ETL |  |  |  |  |  |  | pg 15, 16 |  |  |
|  | quality |  |  |  |  |  |  | pg 26, 31, 47, 48, 49 |  |  |
|  | as a gap affecting interoperability |  |  |  |  |  |  | pg 35, 37, 38, |  | pg 8, 11 |
|  | in data analysis |  |  |  |  |  |  | pg 39, 40 |  |  |
|  | in persistent identifiers [PIDs] |  |  | usecase 6, 25, 32 (error on pg 254) |  |  | pg 42 |  |  |  |
|  | related acronyms |  |  |  |  |  |  | pg 53, 55, 56 |  |  |
|  | related standards |  |  |  |  |  |  | appendix B |  |  |
|  | related to FAIR |  |  |  |  |  |  |  |  | pg 40 |
|  | in open data |  |  |  |  |  |  |  |  | pg 19 |
|  | data checklists |  |  |  |  |  |  |  |  | pg 44, 47, 48, 49, 50, 51, 52, 53, 56, 57, 60, 62, 63 |
|  | metadata issues in mashup |  |  |  |  |  |  |  |  | pg 38 |
|  | related to cleaning |  |  |  |  |  |  |  |  | pg 39 |
|  | in relation to provenance |  |  | sec 8.6 |  |  |  |  |  |  |
|  | usecase 7 (recheck) |  |  | usecase 7 |  |  |  |  |  |  |
|  | usecase 32, 33, and 34 (recheck) |  |  | usecase 32, 33, and 34 |  |  |  |  |  |  |
|  | usecase 45 (recheck) |  |  | usecase 45 |  |  |  |  |  |  |
|  | usecase 2-1 and appendix F (recheck) |  |  | usecase 2-1 and appendix F |  |  |  |  |  |  |
|  | as a key aspect of data science, in template |  |  | sec 3.4 |  |  |  |  |  |  |
|  | types of, and importance to data science | sec 5.4.6, 20 hits |  |  |  |  |  |  |  |  |
|  | definition | sec 2 and sec 3.3.3 |  |  |  |  |  |  |  |  |
|  | data fusion and integration | sec 3.2.3 |  |  |  |  |  |  |  |  |
|  | in the context of the four main characteristics of big data | sec 3.3 |  |  |  |  |  |  |  |  |
|  | in relation to vertical data scientists | sec 5.2 |  |  |  |  |  |  |  |  |
| QUALITY | |  |  |  |  |  |  |  |  |  |
|  | gaps |  |  |  |  |  |  | pg 42, 44 |  |  |
|  | data checklists |  |  |  |  |  |  |  |  | pg 43, 44, 48, 49, 51, 55, 56, 58 |
|  | related to interfaces |  |  |  |  |  |  |  |  | pg 20 |
|  | as an aspect of projects and implementation |  |  |  |  |  |  |  |  | pg 14, 19 |
|  | minimum viable quality |  |  |  |  |  |  |  |  | pg 25 |
|  | levels |  |  |  |  |  |  |  |  | vol 9, p 37 |
|  | FAIR |  |  |  |  |  |  |  |  | pg 40 |
|  | governance |  |  |  |  |  |  |  |  | pg 41 |
| SEARCH | |  |  |  |  |  |  |  |  |  |
|  | applicability to gaps 2 and 4 |  |  |  |  |  |  | pg 37, 38, 39, 40, 42 |  |  |
|  | prevalence in use cases |  |  |  |  |  |  | pg 30, 31, 32 |  |  |
|  | as a data consumer requirement |  |  | sec 3.2 |  |  |  | pg 17 |  |  |
|  | issues in data virtualization |  |  |  |  |  |  | pg 51 |  |  |
|  | related standards |  |  |  |  |  |  | appendix B |  |  |
|  | usecase 15 (recheck) |  |  | usecase 15 |  |  |  |  |  |  |
|  | usecase 21 (recheck) |  |  | usecase 21 |  |  |  |  |  |  |
|  | usecase 2-1 (recheck) |  |  | usecase 2-1 appendix F |  |  |  |  |  |  |
|  | in relation to analytics, template q section 1.35?? |  |  | in relation to analytics, template q section 1.35?? |  |  |  |  |  |  |
|  | usecase 8 (recheck) |  |  | usecase 8, 8 hits |  |  |  |  |  |  |
|  | usecase 2 (recheck) |  |  | usecase 2, 5 hits |  |  |  |  |  |  |
|  | usecase 18 (recheck) |  |  | usecase 18 |  |  |  |  |  |  |
|  | usecase 34 (recheck) |  |  | usecase 34 |  |  |  |  |  |  |

## Thesaurus

This could be a grouping of synonyms for topics. This could be presented in a graphic format. Each reader brings their own set of terminology and understanding. If they are from a particular audience, then a sub topic of interest to them may be obscured due to our use of terminology which is unfamiliar to them.

# Highlights of Select Tables

The following list are some significant tables within the volumes.

* Use Case to requirements mapping in Volume 3
* Use Case to NBDRA mapping (Volume 4, Appendix E)
* NIST Big Data Security and Privacy Safety Levels (Volume 4, Appendix A)

# Additional NIST Resources

Links to other NIST documents and websites that cover these topics more in depth. List will be organized by Volume. Resources outside of NIST will not be included because we don’t have the resources to create a complete list.

Volume 1

Volume 2

Volume 3

Volume 4

* Cloud Security
  + NIST Special Publication 500-322: Evaluation of Cloud Computing Services Based on NIST SP 800-145, February 2018, <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.500-322.pdf>
  + Special Publication 500-291, Version 2: NIST Cloud Computing Standards Roadmap, July2013. <https://www.nist.gov/sites/default/files/documents/itl/cloud/NIST_SP-500-291_Version-2_2013_June18_FINAL.pdf>
  + NIST Special Publication 500-299: DRAFT NIST Cloud Computing 5 Security Reference Architecture, May 2013. <https://collaborate.nist.gov/twiki-cloud-computing/pub/CloudComputing/CloudSecurity/NIST_Security_Reference_Architecture_2013.05.15_v1.0.pdf>
* Risk Management Framework

Volume 5

Volume 6

Volume 7

Volume 8

* Microservices: Microservices, Containers for Vol 8. Draft SP 800-180,

Volume 9

1. Individual Volume Table of Contents

These are the table of contents copied from each volume. They were copied from the drafts on 9/10/19.

* 1. Volume 1, Definitions
  2. Volume 2, Big Data Taxonomies

[Executive Summary vii](#_Toc17199663)

[1 Introduction 1](#_Toc17199664)

[1.1 Background 1](#_Toc17199665)

[1.2 Scope and Objectives of the Definitions and Taxonomies Subgroup 3](#_Toc17199666)

[1.3 Report Production 3](#_Toc17199667)

[1.4 Report Structure 3](#_Toc17199668)

[2 Reference Architecture Taxonomy 6](#_Toc17199669)

[2.1 Actors and Roles 6](#_Toc17199670)

[2.2 System Orchestrator 8](#_Toc17199671)

[2.3 Data Provider 10](#_Toc17199672)

[2.4 Big Data Application Provider 13](#_Toc17199673)

[2.5 Big Data Framework Provider 16](#_Toc17199674)

[2.6 Data Consumer 18](#_Toc17199675)

[2.7 Management Fabric 19](#_Toc17199676)

[2.8 Security and Privacy Fabric 19](#_Toc17199677)

[3 Data Characteristic Hierarchy 20](#_Toc17199678)

[3.1 Data Elements 20](#_Toc17199679)

[3.2 Records 21](#_Toc17199680)

[3.3 Inter-Dependent Records 22](#_Toc17199681)

[3.4 Datasets 23](#_Toc17199682)

[3.5 Multiple Datasets 23](#_Toc17199683)

[4 Summary 25](#_Toc17199684)

[Appendix A: Acronyms 26](#_Toc17199685)

[Appendix B: Bibliography 27](#_Toc17199686)

* 1. Volume 3, Use Cases and General Requirements
  2. Volume 4, Security and Privacy
  3. Volume 5, Architectures White Paper Survey
  4. Volume 6, Reference Architecture
  5. Volume 7, Standards Roadmap
  6. Volume 8, Reference Architecture Interfaces
  7. Volume 9, Adoption and Modernization