NIST Big Data Public Working Group

Breakout 1: Definitions and Taxonomies Discussion Overview

NIST Campus
Gaithersburg, Maryland
June 2, 2017

Breakout 1 – Nancy Grady

Volume 1, Definitions

- Additional Clarification
  - Distinguish fusion from alignment from integrated (variety)
  - Need for retention/deletion policies (governance)
  - Add data sharing discussion (governance)
  - Progression statistics -> data mining -> data science
  - Horizontal vs vertical data scientist (or not a data scientist)
    - Make data scientist an umbrella term?
  - Survey of data scientist definitions
  - Provenance and data citation
- Suggestions
  - Data Acquisition instead of Data Collection

Breakout 1 – Nancy Grady

Volume 1, Definitions (cont.)

- Additions
  - Linked Data and RDF
  - Open Data
  - Expand Machine Learning/Deep Learning (NLP, image, etc)
  - Open Science
  - Data Models (mapping real world)
- Related Topics wrt Big Data (current are mostly infrastructure)
  - Multimedia
  - Streaming data analytics
- Context suggestions - history from digitization
- Pointer – paper “A formal definition of Big Data based on its essential features”

Breakout 1 – Nancy Grady

Volume 2, Big Data Taxonomies

- Expand
  - Metadata at each data level
- Missing
  - Discussion of other Big Data Taxonomies
- Pointers
  - W3C PROV
- Consider
  - Taxonomies vs models (such as layer models)