**NIST Big Data Public Working Group (NBD-PWG)**

**NBD-PWD-2015/M0384**

**Source: NBD-PWG**

**Status: Draft**

**Title: Appendix B Terms and Defs**

**Author:** Ann Racuya-Robbins

Below are a set if definitions to begin our conversation. I have begun on section 3.9.

Appendix B: Terms and Definitions

Big Data consists of extensive datasets\*primarily in the characteristics of volume, variety, velocity, and/or variability that results in new and unprecedented amounts and kinds of value; primarily economic and social\*that requires a governed scalable architecture for the efficient and fair storage, manipulation, analysis and realization of this new value to increase the capability of living, individual social good and the well-being of society as a whole.

The Big Data paradigm consists of the distribution of data systems across horizontally coupled, independent resources to achieve the governed scalability needed for the efficient processing and fair realization of the value inherent in extensive datasets.

Big Data engineering is based on technical paradigms that tend to ignore or remain silent on the societal consequences of Big Data; this is why governance is needed, to guide the technical paradigms to use advanced techniques that not only harness independent resources for building scalable data systems, but also use those advanced techniques to assure the just and fair realization of the societal value inherent in those datasets. Big Data engineering so guided will use advanced techniques that harness the value in independent resources for building governable and governed scalable data systems so that when the characteristics of the datasets require new architectures for efficient, fair storage, manipulation, analysis such architectures will also enable the fair realization of value for the capability of living, individual social good, and the well-being of society as a whole.

Data governance is part of an evolving and dynamic rule set for realizing the societal and economic value from datasets. Data governance involves but is not limited to risk management or administering, or formalizing, discipline (e.g., behavior patterns) around the management of data. Data governance is a reflection of the choices made among normative and competing values and ideals such as-efficiency, economic efficiency-autonomy, individual personal autonomy- distributive justice-corrective justice between the parties-fairness and the like-where parity or equality in bargaining power between the parties is the foremost aspiration.

Value refers to the inherent wealth, economic and social, embedded in any data set that must be governed in order to realize that wealth for all members of the society.