



Big Data Ethics Initiative
Glossary of Terms
DRAFT
 23 April 2015

Term	Description
Application Phase	The phase in big data analytics which puts insights into effect; insight development primarily designed to improve existing processes is part of the application phase.
Beneficial	The value that is concerned with the act of big data analytics creating risks for some individuals (or society as a whole) and benefits for other individuals (or society as a whole); any risks should be counterbalanced by any benefits.
Better Marketing Outcomes for Individuals	Better marketing outcomes for individuals include: better product selection, significant discounts, appropriate linkage to other choices
Better Outcomes for Marketers	Better outcomes for marketers include: improved profitability, enhanced customer relationship, undamaged brand/reputation
Big Data	Things one can do with data at a large scale that cannot be done at a smaller scale to extract new insights or create new forms of value in ways that change markets, organizations, the relationship between citizens and government, and more. ¹
Concept Phase	The phase before any real data analytics takes place; organizations brain storm the reasons for using all the intended data sets, anticipate new data creation, chances for new insights, usefulness of those insights and possibilities of further application.
Contextual Interrogation Worksheet	The assessment, during which the issues are identified that must be resolved, in order to assure an organization's big data project is fair to the full range of stakeholders.
Data	Facts that can be analyzed or used in an effort to gain knowledge or make decisions. ²

¹ Mayer-Schöenberger, V. & K. Cukier (2013), *Big Data: A Revolution That Will Transform How We Live, Work and Think*. Houghton Mifflin Harcourt, New York, p. 6.

² Freed Dictionary

Data Accuracy	Accuracy of data may be directly related to how the data was sourced. Data observed may be more precise than data inferred from an algorithm. Data directly observed may be more accurate than data volunteered by individuals.
Data Origins	The origins or provenance of data include: provided by the individual, scraped from the web, obtained from public sources, provided by third-party aggregator, observed in some fashion, derived from other data (i.e. transformation/manipulation), inferred from analytics, provided by vendor.
Data Transformation	Data from diverse sources in diverse formats are put into a form where the data can be analyzed. The process can impact the accuracy of the data itself and the insights that might come from big data.
Discovery Phase	The phase in big data analytics which yields new insights. In the discovery phase, one does not apply insights, but only conducts the research to illuminate the insights. In the discovery phase, the insights are not usually personally impactful; processes designed purely to create new knowledge are part of the discovery phase.
Durability	The ability to last or endure.
Fair	The value that is concerned with both the impartiality of the insights and applications that are a product of big data analytics, and the application of those insights in a fashion that does not violate laws and norms.
Fairness Issues	Possible fairness issues or repercussions include: regulatory, media, public backlash, discriminatory effects such as economic opportunity, physical security, physical wellbeing, and limiting self-determination.
Half-life	A period of time in which algorithms effectively predict future outcomes.
Individual Risks	Risks to individuals include: physical harm, financial harm, health, reputational, embarrassment, shock or surprise, inappropriate discrimination, misuse of data.
Insight	Understandings of relationships that shed light upon or helps solve a problem.
Integrity	The quality of being honest and having strong moral principles; a trustworthy process requires integrity by all parties.
Linkable Data	Data that is personally identifiable, pseudonymous, device identifiable information is linkable to an individual.
Obligations Associated with Data	The conditions related to data include laws, policies, contracts, industry codes, and individual choices.
Organizational Risks	Risks to organizations include: negative media attention, negative regulatory impact, compliance/reputational, business continuity.
Output	The information produced by big data analytics.

Predictive Value	Performance of a diagnostic test or other statistical measure related to future outcomes.
Pre-processing of Data	The steps in the preparation of data for analytics include: data standardization, data hygiene, and data integration (consolidation).
Progressive	The value that is concerned with the benefit from big data analytics being materially better than not using big data analytics; if the anticipated improvements can be achieved in a less data-intensive manner, then that less intensive approach should be pursued; better outcomes than existing means.
Project Team	Individuals with the following responsibilities should be included on the big data analytics project team: data capture/acquisition, data preparation/management, oversight for restrictions (legal or contractual), and appropriate application of the analysis/insights.
Purpose	The purposes for a big data project include: marketing or risk management, solution and product capability, distribution network, brand enhancement, and marketing (e.g. traditional direct mail, email, telemarketing, digital advertising).
Residual Risk	The risk an activity poses after mitigation strategies have been deployed.
Respectful	The value that is concerned with the context in which the data is originated and the contractual or notice-related restrictions on how the data might be applied.
Reticence	The harm or un-realized value to individuals, organizations or society when data is not used.
Review Phase	The phase in which ongoing assessment occurs in order to assure that controls are working and when new applications of data are introduced.
Risk	The potential of losing something of value or causing harm.
Risk Mitigation	Taking steps to reduce adverse effects.
Stakeholders	Possible stakeholders related to both the analysis and the uses of the resulting insights are: individuals/data subjects (consumers, customers, prospects), organizations (including business and non-governmental organizations), political entities/government, and society/public-at-large/community.
Stakeholder Benefits	Possible stakeholder benefits include: personalization, health, education, economic opportunity, society as a whole, access to new products or services, consumer/customer engagement, brand building, increased revenue, convenience, price.

Stakeholder Concerns	Possible stakeholder concerns that should be taken into account are: data use out of context, contract/legal/privacy /security obligations, data sensitivity, revenue/business needs, cultural differences, commonly held societal values, compatibility with organizational values, and compatibility with social norms regarding the use of sensitive information.
Stakeholder Risks	Possible stakeholder risks include: potential impacts of false positive or negatives.
Structured Data	Relational databases and spreadsheets.
Sustainable	The value that is concerned with how long an insight might endure; the longevity of an outcome affects its sustainability; prolonged longevity of big data analytics is more sustainable; the half-life of an insight affects sustainability.
Synthesized	Different data elements are derived from various source elements.
Transparency and Choice	The relationship between the individual and where the data first originates impacts both transparency and choice. If data is collected by an organization in a direct relationship with the individual, transparency and choice both are direct, and future marketing use may be expected. Future marketing use may be expected if data is collected by a third party for use by the organization with which the individual has a relationship but may not be expected if the data collected by the third party is for use by a different marketer.
Unlinkable Data	Data is not linkable to a person if it is pseudonymous, de-identified, or aggregated.
Unstructured Data	Information that is not organized in a pre-defined manner.