**TBS Data Glossary**

Font Page:

49 BIG DATA TERMS YOU NEED TO KNOW

Big data can be intimidating. Simon Fraser University’s glossary helps you understand the basic concepts of data science so you can feel more confident navigating this field.

Opening:

**SIMON FRASER UNIVERSITY**

As Canada’s engaged university, SFU engages with communities, organizations and partners to create, share and embrace knowledge that improves life and generates real change. We connect research and innovation with government and industry to deliver sustainable and relevant solutions to help solve today’s problems.

* **#1** comprehensive university in Canada. *Maclean’s University Rankings*
* **Top 55** "tech challenger" universities in the world for having innovation at the core of its strategy and strong industry links. *Times Higher Education*
* **Top 4** Computing Science department in Canada. *CSRankings*

**WE LEVERAGE THE POWER OF BIG DATA**

With over a decade of leadership in the big data field, SFU unlocks data to fill a critical talent shortage, increase social impact and contribute to an innovative economy. We empower people to apply the power of big data and artificial intelligence to solve problems that truly matter.

By connecting government, industry and communities together to leverage big data, we are delivering unexpected insight that drives business forward and benefits society.

# Glossary

Learning some of the core concepts and definitions behind big data can help make the field of data science less intimidating. This handy guide clarifies the most important terms in the field. [h3]

### 

### ADVANCED RESEARCH COMPUTING

Advanced research computing (ARC) provides massive computational horsepower and [storage](#Storage-and-Physical-Data-Sources) in a [cloud environment](#CLOUD-COMPUTING) to handle problems and data that are too complex for a single desktop computer. [[1]](#endnote-1),[[2]](#endnote-2)

### ALGORITHM

A sequence of instructions telling a computer how to answer a specific question. [[3]](#endnote-3),[[4]](#endnote-4)

### ANALYTICS

The process of using [statistical models](#MACHINE-LEARNING) and software to [transform data](#DATA-TRANSFORMATION) into useful [information](#INFORMATION) and to draw conclusions towards effective decision making.[[5]](#endnote-5),[[6]](#endnote-6)

### ARTIFICIAL INTELLIGENCE

A branch of computer science that allows machines to acquire and apply knowledge to handle new inputs and analyze patterns to solve different problems. [[7]](#endnote-7)

### BEHAVIOUR(AL) ANALYTICS

A type of business [analytics](#ANALYTICS) that examines behavioural data about people to understand how and why individuals act the way they do and to make more accurate predictions for future behaviour.[[8]](#endnote-8),[[9]](#endnote-9)

### BIG DATA

Refers to the massive amounts of [data](#DATA) generated around the world that is too large, complex or varied for traditional processing software. Its potential to be analyzed for valuable information is enabled by technology such as [advanced research computing](#ADVANCED-RESEARCH-COMPUTING). [[10]](#endnote-10),[[11]](#endnote-11)

### CAUSAL INFERENCE

Determine whether observations made in one variable are the reason for an effect observed in another variable, possibly occurring at a later time. See also [correlation mining](#CORRELATION-MINING) and [predictive analytics](#PREDICTIVE-ANALYTICS).

### CLASSIFICATION

An approach in [machine learning](#MACHINE-LEARNING) where a program is [trained](#MODEL-FITTING) with labeled [data](#DATA) to determine which category a new observation belongs to. See also, [clustering](#CLUSTERING).

### CLOUD COMPUTING

Access to data, [storage](#DISTRIBUTED-FILE-SYSTEM), applications, and other computing resources made available to many users on-demand over the [Internet](#NETWORK) to improve [scalable computing power](#ADVANCED-RESEARCH-COMPUTING) and reliability. [[12]](#endnote-12),[[13]](#endnote-13)

### CLUSTER

A [network](#NETWORK) of computers (or compute nodes) that work on tasks together in parallel. The concept is different from [clustering](#CLUSTERING).

### CLUSTERING

A [machine learning](#MACHINE-LEARNING) technique that groups similar data points together to uncover hidden structure. As opposed to [classification](#CLASSIFICATION), clustering does not require labelled training data and is a type of unsupervised learning.

### COMPUTING WITH DATA

Learning from data using computing tools and programming languages, such as Python or R with large ecosystems of libraries providing a convenient level of abstraction and statistical methods that can be organized into data processing pipelines.[[14]](#endnote-14)

### CORRELATION MINING

As a measure of association between two variables, correlation can be used for [prediction](#PREDICTIVE-ANALYTICS) and indicate the presence of [causal relationship](#CAUSAL-INFERENCE). It does, however, not sufficiently imply causation. [[15]](#endnote-15)

### CYBERSECURITY

See [Secure Computing](#SECURE-COMPUTING).

### DATA

Collection of examples, observations, measurements, facts, points, or other items of information that can be represented in [structured](#STRUCTURED-DATA) or [unstructured](#UNSTRUCTURED-DATA) form.

### DATABASE

Organized collection of [data](#DATA) that allows easy access, management, updating and analysis of data. Commonly used databases are MySQL, PostgreSQL, as well as various [NoSQL](#NO-SQL-DATABASE) options.[[16]](#endnote-16)

### DATA EXPLORATION AND PREPARATION

Exploratory data analysis (EDA) is a formative step in the [creation of models](#MODEL-FITTING). [Views of the data](#VISUALIZATION) are used to learn about patterns or relationships among variables. This includes data cleaning (careful removal of erroneous or unreliable data points) and manipulation for further analysis.

### DATA INTEGRATION

The process of combining information from different data sources in preparation for data processing. [[17]](#endnote-17)

### DATA MINING

An analytical process where large datasets are explored or "mined" in search of meaningful patterns, relationships or insights. The process can include statistics, [machine learning](#MACHINE-LEARNING) or other forms of artificial intelligence.[[18]](#endnote-18),[[19]](#endnote-19)

### DATA REPRESENTATION

The form in which [data](#DATA) is stored, processed and transmitted such that its [information content](#INFORMATION) and [context](#METADATA) are retained as much as possible. Choices of form are influenced by hardware, software or other constraints around processing and analysis resources. [[20]](#endnote-20)

## DATA TRANSFORMATION

The process of converting data from [one form into another](#DATA-REPRESENTATION) to gain better [insight](#INSIGHT), for instance, converting [unstructured data](#UNSTRUCTURED-DATA) into [structured form](#STRUCTURED-DATA) that can be analyzed further.

### DATA SCIENCE

A multidisciplinary activity combining programming skills, math and statistical analysis, and sector-specific expertise to extract [insights](#INSIGHT) from [data](#DATA). Often performed in stages: 1. [Data Exploration and Preparation](#DATA-EXPLORATION-AND-PREPARATION), 2. [Data Representation](#DATA-REPRESENTATION) and [Transformation](#DATA-TRANSFORMATION), 3. [Computing with Data](#COMPUTING-WITH-DATA), 4. [Machine Learning](#MACHINE-LEARNING), 5. [Data Visualization and Presentation](#VISUALIZATION). [[21]](#endnote-21),[[22]](#endnote-22)

### DESCRIPTIVE ANALYTICS

An initial stage of data processing that involves creating a summary of historical data with the goal of answering the question, "What happened?"[[23]](#endnote-23),[[24]](#endnote-24)

### DISTRIBUTED FILE SYSTEM

A mechanism that stores files on servers and allows clients, with permission, to store and process files as if they were stored on their own computer.[[25]](#endnote-25),[[26]](#endnote-26)

### FEATURE

A key property that characterizes a data point (representing some real-world object) in the context of a [machine learning](#MACHINE-LEARNING) problem or other type of [analysis](#ANALYTICS).

### FEATURE ENGINEERING

A human-driven process of finding the most important [features](#FEATURE) to develop [predictive](#PREDICTIVE-ANALYTICS) models. Formerly, as "art" to the "science" of [machine learning](#MACHINE-LEARNING), [big data](#BIG-DATA) increasingly moves towards automated representation learning.

### FLEXIBLE SCHEMA

Unlike [SQL](#SQL) databases with a [tabular](#TABULAR-DATA) [structure](#STRUCTURED-DATA), objects or documents stored in a flexible schema can be different from one another.[[27]](#endnote-27)

### GENERATIVE MODELING

[Fitting](#MODEL-FITTING) a [model](#MACHINE-LEARNING) that can generate synthetic data beyond [given observations](#DATA). For example, a language model trained on a collection of text can be used to suggest possible words to modify or continue a given phrase. [[28]](#endnote-28)

### GOODNESS OF FIT

Summarizes how well the values observed in the data agree with those values expected by the model.

### INFORMATION

Meaning encoded in [data](#DATA) that answers questions to better understand a concept by interpreting data within the context of its problem setting or domain. See also data analysis [algorithms](#ALGORITHM) and information [visualization](#VISUALIZATION). [[29]](#endnote-29)

### INSIGHT

Actionable [information](#INFORMATION) gained by interpreting [data analysis](#ANALYTICS) results, helping people to make more informed decisions. [[30]](#endnote-30)

### IOT (INTERNET OF THINGS)

A system of [connected](#NETWORK), "smart" objects, including smartphones, wearables and smart-appliances, that collect and exchange information without requiring human interaction. The IoT is a huge generator of data.[[31]](#endnote-31),[[32]](#endnote-32)

### KNOWLEDGE DISCOVERY

Knowledge discovery (KDD) aims to extract [insight](#INSIGHT) from data in [databases](#DATABASE). It involves a [number of steps](#DATA-SCIENCE) including the evaluation and possibly [interpretation](#VISUALIZATION) of [patterns](#DATA-MINING) to gather [insight](#INSIGHT) and knowledge.[[33]](#endnote-33)

### MACHINE LEARNING

A part of [artificial intelligence](#ARTIFICIAL-INTELLIGENCE) that enables machines to learn from experience to perform certain tasks by using [algorithmic](#ALGORITHM) models that are [trained](#MODEL-FITTING) to imitate patterns present in [data](#DATA) to support [prediction](#PREDICTIVE-ANALYTICS), [data generation](#GENERATIVE-MODELING) and other forms of [data analytics](#ANALYTICS).

### METADATA

Also called "data about data," it provides a [structured description](#STRUCTURED-DATA) and context for a data point (document, image, or file) to help organize, find and understand the data. [[34]](#endnote-34)

### MODEL FITTING

Optimizes [how well](#GOODNESS-OF-FIT) a [machine learning model](#MACHINE-LEARNING) can make predictions for previously unseen test data.

### NETWORK

A system where multiple computing devices are connected to each other to exchange information and resources through a data link. For example, the Internet.[[35]](#endnote-35),[[36]](#endnote-36)

### NETWORK ANALYSIS

Mapping and measuring the relationships between people, groups, organizations, computers and other connected entities. It is used to simplify complex relationships, making them easier to analyze. [[37]](#endnote-37)

### NO-SQL DATABASE

A type of database designed to handle [large](#ADVANCED-RESEARCH-COMPUTING) volumes of data that [may not have a structure](#FLEXIBLE-SCHEMA).

### PREDICTIVE ANALYTICS

A process for analyzing current [data](#DATA) to determine future events or other unknowns. Related to [descriptive](#DESCRIPTIVE-ANALYTICS) and [prescriptive analytics](#PRESCRIPTIVE-ANALYTICS), it draws on techniques from [data mining](#DATA-MINING), modeling, [machine learning](#MACHINE-LEARNING) and statistics.[[38]](#endnote-38),[[39]](#endnote-39)

### PRESCRIPTIVE ANALYTICS

A type of [analytics](#ANALYTICS) with the goal of using data to determine the best course of action for a specific scenario. [[40]](#endnote-40)

### REGRESSION

A type of [statistical model](#MACHINE-LEARNING) that [predicts](#PREDICTIVE-ANALYTICS) numerical values (instead of [class labels](#CLASSIFICATION)).

### RELATIONSHIP MINING

Relationship mining examines associations between two or more variables in a dataset, for example, by [correlation mining](#CORRELATION-MINING) and [causal mining](#CAUSAL-INFERENCE).

### SECURE COMPUTING

Efforts to ensure privacy and to protect data, devices and computing systems from harm (such as hacking, damage, and malpractice), and mitigate service disruptions. [[41]](#endnote-41),[[42]](#endnote-42)

### SQL

SQL is a language that works with structured, [tabular data](#TABULAR-DATA) allowing to query and construct such data. It forms a standard for many [database](#DATABASE) systems. [28]

### STRUCTURED DATA

[Data](#DATA) that is organized into clearly defined fields associated with variables or attributes, such as dates, words, or numbers that are recorded for each observation or item. Items are often represented as rows of [tabular spreadsheets](#TABULAR-DATA) that can be stored in a [database](#DATABASE) for easy processing and analysis.[[43]](#endnote-43),[[44]](#endnote-44)

### TABULAR DATA

Data items or observations that are organized as rows that contain values under columns that correspond to specific variables or properties. [[45]](#endnote-45),[[46]](#endnote-46)

### TEXT MINING

Text mining is a process of analyzing [text](#UNSTRUCTURED-DATA) to capture key concepts, themes, relationships and trends.[[47]](#endnote-47)

### UNSTRUCTURED DATA

Data that is not organized in a pre-defined way under a single data model. Examples include text, images, audio, or video. [[48]](#endnote-48),[[49]](#endnote-49)

### VISUALIZATION

Visual representation of data including plots, charts, maps and infographics to support people with tasks such as sensemaking, effective communication of information, and improved pattern detection. For example, flight operation using a radar screen that displays sensor data in real-time. [[50]](#endnote-50),[[51]](#endnote-51)

### VISUALIZATION DASHBOARD

Multiple interactive graphical views of [data](#DATA) that would be too complex if presented in a [single visualization](#VISUALIZATION), for instance, to support complex narratives and [insights](#INSIGHT) by showing key performance indicators (KPIs) for monitoring of a business or other processes.

# Categories

The categories offer an alternative grouping that helps to find related terms.

### Data Inquiry Disciplines

[ADVANCED RESEARCH COMPUTING](#ADVANCED-RESEARCH-COMPUTING)  
[ANALYTICS](#ANALYTICS)  
[BIG DATA](#BIG-DATA)  
[CLOUD COMPUTING](#CLOUD-COMPUTING)  
[COMPUTING WITH DATA](#COMPUTING-WITH-DATA)  
[CYBERSECURITY](#SECURE-COMPUTING)  
[DATA MINING](#DATA-MINING)  
[DATA SCIENCE](#DATA-SCIENCE)  
[KNOWLEDGE DISCOVERY](#KNOWLEDGE-DISCOVERY)  
[MACHINE LEARNING](#MACHINE-LEARNING)  
[SECURE COMPUTING](#SECURE-COMPUTING)

### Data Analysis Techniques

[BEHAVIOUR(AL) ANALYTICS](#BEHAVIOURAL-ANALYTICS)  
[CAUSAL INFERENCE](#CAUSAL-INFERENCE)  
[CLASSIFICATION](#CLASSIFICATION)  
[CLUSTERING](#CLUSTERING)  
[CORRELATION MINING](#CORRELATION-MINING)  
[DESCRIPTIVE ANALYTICS](#DESCRIPTIVE-ANALYTICS)  
[FEATURE ENGINEERING](#FEATURE-ENGINEERING)  
[GENERATIVE MODELING](#GENERATIVE-MODELING)  
[NETWORK](#NETWORK)  
[NETWORK ANALYSIS](#NETWORK-ANALYSIS)  
[PREDICTIVE ANALYTICS](#PREDICTIVE-ANALYTICS)  
[PRESCRIPTIVE ANALYTICS](#PRESCRIPTIVE-ANALYTICS)  
[REGRESSION](#REGRESSION)  
[RELATIONSHIP MINING](#RELATIONSHIP-MINING)  
[TEXT MINING](#TEXT-MINING)  
[VISUALIZATION](#VISUALIZATION)  
[VISUALIZATION DASHBOARD](#VISUALIZATION-DASHBOARD)

### Data Processing

[ALGORITHM](#ALGORITHM)  
[DATA AGGREGATION](#DATA-INTEGRATION)  
[DATA EXPLORATION AND PREPARATION](#DATA-EXPLORATION-AND-PREPARATION)  
[DATA REPRESENTATION](#DATA-REPRESENTATION)  
[DATA TRANSFORMATION](#DATA-TRANSFORMATION)  
[GOODNESS OF FIT](#GOODNESS-OF-FIT)  
[SQL](#SQL)

### Storage and Physical Data Sources

[DATABASE](#DATABASE)  
[DISTRIBUTED FILE SYSTEM](#DISTRIBUTED-FILE-SYSTEM)  
[IOT (INTERNET OF THINGS)](#IOT-INTERNET-OF-THINGS)  
[NoSQL](#NO-SQL-DATABASE)

### Types of Data

[FLEXIBLE SCHEMA](#FLEXIBLE-SCHEMA)  
[METADATA](#METADATA)  
[STRUCTURED DATA](#STRUCTURED-DATA)  
[TABULAR DATA](#TABULAR-DATA)  
[UNSTRUCTURED DATA](#UNSTRUCTURED-DATA)

Closing:

**ENGAGE WITH SFU**

Simon Fraser University is empowering people to unlock data for insights and impact.Engage a trusted partner to enable better decision-making that transforms our society and economy.

* Explore how big data can improve your organization, services and processes.
* Collaborate with leading experts to achieve impactful policy outcomes using tailored data-driven approaches.
* Build the technical and strategic skills your team needs to make better data-driven decisions.
* Discover the next generation of job-ready big data leaders applying new and cutting-edge approaches.

Back page:

[bottom half]

CANADA’S ENGAGED UNIVERSITY  
Start your data journey today: sfu.ca/big-data

# References

1. "Westgrid: What We Do." [Online]. Available: <https://www.westgrid.ca/about_westgrid/what_we_do> [↑](#endnote-ref-1)
2. "Compute Canada. Update January 11, 2017." [Online]. Available: <https://www.computecanada.ca/wp-content/uploads/2015/02/Update-January-11-2017.pdf> [↑](#endnote-ref-2)
3. "A beginner's guide to big data terminology." [Online]. Available: <http://dataconomy.com/2016/05/a-beginners-guide-to-big-data-terminology> [↑](#endnote-ref-3)
4. "Mini-glossary: Big data terms you should know." [Online]. Available: <http://www.techrepublic.com/article/mini-glossary-big-data-terms-you-should-know> [↑](#endnote-ref-4)
5. "A beginner's guide to big data terminology." [Online]. Available: <http://dataconomy.com/2016/05/a-beginners-guide-to-big-data-terminology> [↑](#endnote-ref-5)
6. "Mini-glossary: Big data terms you should know." [Online]. Available: <http://www.techrepublic.com/article/mini-glossary-big-data-terms-you-should-know> [↑](#endnote-ref-6)
7. "Artificial Intelligence". Available: https://www.techopedia.com/definition/190/artificial-intelligence-ai [↑](#endnote-ref-7)
8. "Behavioral analytics, definition by Techopedia." [Online]. Available: <https://www.techopedia.com/definition/30308/behavioral-analytics> [↑](#endnote-ref-8)
9. "Detecting advanced threats with user behavior analytics." [Online]. Available: <http://www.networkworld.com/article/2904356/security0/detecting-advanced-threats-with-user-behavior-analytics.html> [↑](#endnote-ref-9)
10. "Big data definition by Tech Target." [Online]. Available: <http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data> [↑](#endnote-ref-10)
11. "What Is Big Data?" [Online]. Available: <https://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/#4fb6509b5c85> [↑](#endnote-ref-11)
12. "Who Coined 'Cloud Computing'?" [Online]. Available: <https://www.technologyreview.com/s/425970/who-coined-cloud-computing> [↑](#endnote-ref-12)
13. "Cloud computing: a simple introduction." [Online]. Available: <http://www.explainthatstuff.com/cloud-computing-introduction.html> [↑](#endnote-ref-13)
14. Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745-766. [↑](#endnote-ref-14)
15. "Introduction to Correlation". Available: https://www.datascience.com/blog/introduction-to-correlation-learn-data-science-tutorials [↑](#endnote-ref-15)
16. "A beginner's guide to big data terminology." [Online]. Available: <http://dataconomy.com/2016/05/a-beginners-guide-to-big-data-terminology> [↑](#endnote-ref-16)
17. "Mini-glossary: Big data terms you should know." [Online]. Available: <http://www.techrepublic.com/article/mini-glossary-big-data-terms-you-should-know> [↑](#endnote-ref-17)
18. "A beginner's guide to big data terminology." [Online]. Available: <http://dataconomy.com/2016/05/a-beginners-guide-to-big-data-terminology> [↑](#endnote-ref-18)
19. "Mini-glossary: Big data terms you should know." [Online]. Available: <http://www.techrepublic.com/article/mini-glossary-big-data-terms-you-should-know> [↑](#endnote-ref-19)
20. Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745-766. [↑](#endnote-ref-20)
21. Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745-766. [↑](#endnote-ref-21)
22. "The Data Science Venn Diagram". Available: <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram> [↑](#endnote-ref-22)
23. "Descriptive analytics definition by Tech Target." [Online]. Available: <http://whatis.techtarget.com/definition/descriptive-analytics> [↑](#endnote-ref-23)
24. "Four Types of Big Data Analytics and Examples of Their Use." [Online]. Available: <http://www.ingrammicroadvisor.com/data-center/four-types-of-big-data-analytics-and-examples-of-their-use> [↑](#endnote-ref-24)
25. "Distributed file system definition by Tech Target." [Online]. Available: <http://searchwindowsserver.techtarget.com/definition/distributed-file-system-DFS> [↑](#endnote-ref-25)
26. "Distributed file system definition by Techopedia." [Online]. Available: <https://www.techopedia.com/definition/1825/distributed-file-system-dfs> [↑](#endnote-ref-26)
27. "Data Modeling Introduction". Available: <https://docs.mongodb.com/manual/core/data-modeling-introduction/> [↑](#endnote-ref-27)
28. Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745-766. [↑](#endnote-ref-28)
29. "Data vs. Information vs. Insight". Available: https://online.ben.edu/programs/mba/resources/data-vs-information-vs-insight [↑](#endnote-ref-29)
30. "Data vs. Information vs. Insight". Available: https://online.ben.edu/programs/mba/resources/data-vs-information-vs-insight [↑](#endnote-ref-30)
31. "A Simple Explanation Of 'The Internet Of Things'." [Online]. Available: <https://www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/#34e4f89b1d09> [↑](#endnote-ref-31)
32. "Internet of things definition by Tech Target." [Online]. Available: <http://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT> [↑](#endnote-ref-32)
33. "Overview of the KDD Process". Available: http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1\_kdd.html [↑](#endnote-ref-33)
34. Greenberg, J. (2003). Metadata and the world wide web. *Encyclopedia of library and information science*, *3*, 1876-1888. [↑](#endnote-ref-34)
35. "Computer network definition by Techopedia." [Online]. Available: <https://www.techopedia.com/definition/25597/computer-network> [↑](#endnote-ref-35)
36. "What is computer networking?" [Online]. Available: <https://www.lifewire.com/what-is-computer-networking-816249) [www.lifewire.com/what-is-computer-networking-816249.](https://www.lifewire.com/what-is-computer-networking-816249> [↑](#endnote-ref-36)
37. "Social Network Analysis: An Introduction by Orgnet,LLC." [Online]. Available: <http://www.orgnet.com/sna.html> [↑](#endnote-ref-37)
38. "A beginner's guide to big data terminology." [Online]. Available: <http://dataconomy.com/2016/05/a-beginners-guide-to-big-data-terminology> [↑](#endnote-ref-38)
39. Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745-766. [↑](#endnote-ref-39)
40. "A beginner's guide to big data terminology." [Online]. Available: <http://dataconomy.com/2016/05/a-beginners-guide-to-big-data-terminology> [↑](#endnote-ref-40)
41. "The myth of secure computing." [Online]. Available: <https://hbr.org/2003/06/the-myth-of-secure-computing) [hbr.org/2003/06/the-myth-of-secure-computing.](https://hbr.org/2003/06/the-myth-of-secure-computing> [↑](#endnote-ref-41)
42. "Mini-glossary: Big data terms you should know." [Online]. Available: <http://www.techrepublic.com/article/mini-glossary-big-data-terms-you-should-know> [↑](#endnote-ref-42)
43. "Introduction to Structured Data, by Google." [Online]. Available: <https://developers.google.com/search/docs/guides/intro-structured-data> [↑](#endnote-ref-43)
44. "Structured data definition by Webopedia." [Online]. Available: <http://www.webopedia.com/TERM/S/structured_data.html> [↑](#endnote-ref-44)
45. "Introduction to Tabular Data" Available: <https://papl.cs.brown.edu/2016/intro-tabular-data.html> [↑](#endnote-ref-45)
46. "Tabular Database". Available: <https://www.techopedia.com/definition/26181/tabular-database> [↑](#endnote-ref-46)
47. "About Text Mining". Available: https://www.ibm.com/support/knowledgecenter/en/SS3RA7\_15.0.0/com.ibm.spss.ta.help/tm\_intro\_tm\_defined.htm [↑](#endnote-ref-47)
48. "Unstructured data definition by Techopedia." [Online]. Available: <https://www.techopedia.com/definition/13865/unstructured-data> [↑](#endnote-ref-48)
49. "Solving the Unstructured Data Challenge." [Online]. Available: <http://www.cio.com/article/2941015/big-data/solving-the-unstructured-data-challenge.html> [↑](#endnote-ref-49)
50. "Data visualization definition by Tech Target." [Online]. Available: <http://searchbusinessanalytics.techtarget.com/definition/data-visualization> [↑](#endnote-ref-50)
51. "Data visualization for human perception." [Online]. Available: <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception> [↑](#endnote-ref-51)