Towards Developing Theories About Data: A Philosophical and Scientific Approach

Rahmatollah Fattahi

Department of Library and Information Science, Ferdowsi University of Mashhad, Iran,
<fattahirahmat@gmail.com>

Rahmatollah Fattahi holds a BA in English Language and literature, and MS and PhD in Library and Information Sciences. He is a Full Professor at the Department of Library and Information Sciences, Ferdowsi University of Mashhad, Iran. Fattahi's research interests cover Knowledge Organization, Information Retrieval, and Human-Computer Interaction. He has published a number of Persian books on Information Organization and also English papers in international journals and has so far attended a number of international conferences. Fattahi was the president of the Iranian Library and Information Science Association during 2000-2003 and 2006-2009. Also he has been the chair of the ISKO Iranian Chapter for about 10 years.

Fattahi, Rahmatollah. 2022. "Towards Developing Theories About Data: A Philosophical and Scientific Approach." *Knowledge Organization* 49(3): 141-150. 56 references. DOI:10.5771/0943-7444-2022-3-141.

Abstract: Similar to information and knowledge, data and especially big data are now known as one of the most vital elements in the 21st century since they provide multiple capabilities to individuals and organizations. However, in comparison to some theories about information and knowledge, there are no significant attempts in most scientific disciplines for building theories about data. This paper first reviews the different definitions provided about the concept of data in the works of scholars. It then identifies and explores the philosophical aspects as well as the multiple capabilities/features that can be derived from data. Finally, a starter list of some basic/general theories is developed based on the capabilities and features of data. Such new theories can be used as meta-theories to extend data theories for various scientific disciplines. The important notion supporting the development of theories about data is that, if data is so important and if data science is to continue flourishing in a variety of specialized fields and trends, then we need to build relevant theories about data for research and practical purposes in a multi-disciplinary context.

Received: 29 November 2020; Revised 4 July 2022; Accepted 18 July 2022

Keywords: Data, big data, theory of data, philosophy of data.

1.0 Introduction

Data, information and knowledge are three important associated concepts for the understanding and development of modern sciences. Among the three concepts, data has recently gained more attention by scholars since it provides fundamental, scientific evidence about the environment, living systems and processes. Studying data helps us achieve a deeper understanding of ecosystems, entities[1] and the relationships between and/or among them. Such an understanding can lead us to the further flourishing of data science (see Gil 2013 for a short history and timeline; Van der Aalst 2016; Donoho 2017; Cao 2017), as well as the development of new educational and research trends in many scientific disciplines.

The present paper seeks to argue that understanding the ecosystems in the real world, their structure and their constituent elements, as well as the events,[2] requires an understanding of both theoretical applied aspects of the data. The term data is now used frequently in many fields including Library and Information Science (Hjørland 2018, 685) and Computer Sciences. Development of the theoretical foundations of modern sciences such as computer science, data science, neuroscience, and other related sciences is, to a great extent, dependent on the understanding of the real value of data and formulation of relevant theories about it.

2.0 The value of data for humans and organizations

It is only in the last few decades that human beings have realized the value of data as a crucial element which has extensive capabilities for individual, social, or organizational life (Mons et al. 2011; Vertesi and Douris 2011; Labrinidis and Jagadish 2012; Carnahan 2014; Pouchard et al. 2015; Maycotte 2015; Hakeem ;2016 Vandeventer 2016).

Similar to information and knowledge, data and especially big data is now known as one of the most vital elements in the 21st century since it provides multiple capabilities to individuals and organizations. There are considerable references in the literature emphasizing the value of data. For example, Undavia, Patel and Shah (2012) believe that data is the most vital part of any business or organization; Shahzad, Rehman and Qamar (2017), Shweta and Muralidhara (2018) and Dessaei et. al. (2019) consider data as the most vital element that must be treated as an asset.

On the other hand, and from the neuroscience perspective, findings have shown us how data is captured and stored by the nervous system and becomes the main element for living systems to function .Similar to other animals, human beings are always in interaction with the environment (i.e., ecosystems covering entities, relationships and evidences). They perceive entities, phenomena and evidences through their five senses and process the signals they capture through their nervous system. This is done by receiving signals from entities and phenomena presented in the outside world through receptors of the five sensory organs in the form of electromagnetic waves (sight), sound waves (hearing), airborne particles (smell), chemical molecules (taste), and/or mechanical and thermal changes (touch). Such processes occur in the nervous system based on the data representation of entities and phenomena. In other words, to understand, process, store, and recall what is captured through the nervous system, sensory receptors convert each of these variables into electrochemical messages for coding and further processing (Baars and Gage 2013; Atkinson, Atkinson, Smith, Bem and Nolen-Hoeksema 2000). This is a basic requirement for forming information in the mind. The most important data processing centres in the brain are the synapses. Capabilities such as memory and learning appear to be based on the characteristics of synapses and the relationship between nerve cells in human beings (Baars and Gage 2013).

The result of such processes is a complex set of reactions to stimuli. Sensory and mental experiences are stored in the form of data in different parts of the nervous system (Baars and Gage 2013) to be used later in various processes during life and interaction with the environment. In this context data is the basic element for mentally processing of the representation of the real world and, when needed, turn that into information and knowledge. In any mental process, signals can potentially be captured as data, information, or knowledge. Data, information, or knowledge, as will be discussed later, is the name given to what shapes our different mental states at the moment of interaction with or observation of entities and events. These observations are either facts recorded, things observed, etc. and once internalized or processed by a human being they can be used for understanding and decision making.

To better understand the context for discussing the concept of data, first we need to deal with the philosophy of data.

3.0 Philosophy of data

There is little reference in the existing literature to the philosophy of data because of the limited theoretical inquiries and discussions published so far. The whats and whys of this crucial element is still not much known and has not been well established scientifically. In the last few decades when the concept of data has come to the scientific field, little attempt has been made to philosophically explain the term beyond its classical meaning. It is logical that, if the value and significance of an element, concept, or phenomenon in the universe and in human life is constantly growing, then we must delve into its whats and whys (i.e., its philosophy). Emphasizing the need for a philosophy of data, Furner (2017, 55) states that:

> Philosophy of data should not be dismissed as a cluster of scholastic puzzles whose solutions are of limited practical value. On the contrary, philosophy of data should be recognized as constituting the core of a field of data studies that is informed by, but far from equivalent to, statistics, computer science, and library and information studies.

He points out that the philosophy of data should be the core of studies in mathematics, computer science, librarianship, and information science. Furner (2017) cites two articles by Ballsun-Stanton and Bunker (2009) and Leonelli (2016), which are among the few philosophical writings on data. Furner (2017) also reports Franks (1966) as probably the first to talk about the philosophy of data. He points to the fact that Franks used the label “philosophy of data” to refer to what we might nowadays call a “data model”. As Furner (55) believes, Franks certainly was not talking about “a field of inquiry, like philosophy of science or philosophy of language”.

To philosophically justify the value of data, first it is important to re-emphasize that, data are the core elements of life in the real world. Today, much of the activities of many individuals and organizations are substantially based on the capturing, production, distribution, processing, and consumption of data. In other words, activities such as the production, distribution, and consumption of data are among the important indicators of dynamic and advanced societies. Floridi (2008, 236) emphasizes that “the more societies move towards data-driven societies, the more concerned and careful they need to become about their very foundation.”

According to Ballsun-Stanton and Bunker (2009, 123), the existential philosophy of data can be explained through three distinct aspects:

1. the first aspect is the "technological understanding" of data as bits; that is, the data element is encoded with digital information;
2. the second aspect is the "scientific understanding" of data as hard numbers, which is itself the result of objective and reproducible measurements;
3. the third aspect emphasizes the understanding of data engineering as the elements of observation, that is, it is the result of subjective perceptions (123).

Furner (2017) is one of the few scholars addressing this issue. He points to the historical roots of the use of the term "philosophy of data". He (57) outlines the philosophy of data in three overlapping branches or fields:

1. The ontology (or metaphysics) of data: What, precisely, are data? Of what kind (or genus) of thing are they? Under what conditions can something count as a datum? What properties must something have if it is to be, or perhaps to play the role of, a datum? While it may seem arcane, ontology of data is potentially of great practical use.
2. Epistemology of data: What kinds of knowledge can we have about data (or about the concept data)? In what ways may we acquire or produce that knowledge? Epistemological questions and answers are useful to the extent that an understanding of the practices by which we try to find out about certain phenomena (in this case, the phenomenon of data) allows us to evaluate those practices and potentially develop new, better ones.
3. Ethics and politics of data: What kinds of value do we, could we, and should we place in data and data practices of various kinds? Awareness of the possibilities in these respects allows us to evaluate the phenomenon of data in ways that best support our policymaking goals. What are the social impacts of the collection, manipulation, and distribution, etc., of data?

Besides a few papers explaining the philosophy of data, some authors have also attempted to clarify the philosophy of big data. Swan (2015) believes that philosophy of big data is evolving into a discipline at two levels, i.e., internal and external. It is not only "a generalized articulation of the concepts, theory, and systems that comprise the overall conduct of big data science" but also as "a consideration of the impact of big data science more broadly on individuals, society, and the world". Sun and Strang (2018) propose the general philosophical principles as well as the computational, cognitive, social and economic philosophy of big data. Horne (2018) talks about big data from a complexity standpoint especially in the context of artificial brains and minds. Choenni, Netten, Bargh and Choenni, (2018, 71) argue the philosophical view of big data and discuss the "fundamental objections on the straightforward use of Big Data outcomes".

In sum, understanding the philosophical aspects of data largely depends on the understanding of the definitions that have been presented from different perspectives. Therefore, we need to review various definitions of the concept of data.

4.0 An overview of data definitions: similarities and ambiguities

Philosophical, historical, and comparative analysis of the concept of data illustrates its various aspects from different perspectives leading to different definitions. As Capurro and Hjørland (2003) emphasize, defining terms and concepts in information science is a fundamental issue. An overview of the different definitions of data in the literature indicates that there is still no consensus in this regard. Even the word data has different meanings for different people and is in a state of confusion. Carrying out a critical Delphi study, Zins (2007) reviewed this variety of definitions and pointed out that the overview of such definitions suggests different approaches to the concept of data. Similarly, Borgman, Wallis and Enyedy (2007) and Borgman (2010) point to the differences in the interpretation of data by scientists and their partners in computer science and engineering. Furner (2016, 287) believes that: "Oddly, perhaps, for such a basic term, 'data' has not been as frequently subject to probing analysis in the scholarly literature as 'information'."

Various attempts have so far been made by different authors to define data. Some definitions have indirectly given data a reflection of the facts in the mind (Dewey 2019), the most important of which is the dependence on our understanding that can be logically different or even incomplete and false. The same difference in perceptions and mindsets perceive data primarily as an abstract concept, meaning that we, the human beings, assign meaning to data based on our mental schemas and the context in which we are thinking. Once these meanings are understood and recorded in our minds to create a connection, then the data becomes objective and a reality understandable by others (we call this a 'fact'). William (2014) believes that data is volatile. In other words, events (for example, climate change, storms, etc.) always occur and produce signals, but they cannot be recorded and stored by themselves; they require human beings or machines' intervention to capture, store, and process. When data can be measured, processed, and exchanged, it is received and stored in the form of intelligible facts by human beings or machines. As an example, we receive heat intensity through a sense of touch (skin), send it to our nervous system and perceive it in our mind, but at the same time we have a device called a thermometer that shows the degree of heat with a number (we call this 'data').

The Oxford English Dictionary provides the following definition of data from two perspectives:

> The quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media... Things known or assumed as facts, making the basis of reasoning or calculation.

As mentioned earlier, the diverse set of definitions of what data looks like are related to different people's perspectives or to different contexts. Some authors refer to data as sensory stimuli (Zins 2007 quoting from Shifra Baruchson–Arbib in his Delphi study), or as symbols and raw materials (Abiteboul 1997; Torra 2003; Berlinger et al. 2004; Ackoff 1989); others see it as a set of facts (William 2014), and some scholars even consider data as a replacement for concepts, meanings, or representations of entities (Choo 1996). Rowley (2007, quoted in Hjørland 2018, 686) considers data as "discrete, objective facts or observations". Hjørland (2018) has comprehensively and critically reviewed the many definitions of data provided by different scholars. He (685) states that "epistemologically it is important to establish that what is considered data by somebody need not be data for somebody else." He also criticizes the different definitions set forward by different authors and suggests that a new definition is needed with epistemological understanding. For example, he questions the definition provided by Furner (2017, 66; quoted in Hjørland 2018, 690):

> "data are concrete instantiations of symbolic representations of descriptive propositions, informed by empirical observation, about the quantitative and qualitative properties of real-world phenomena". Undoubtedly, this definition covers most of what should be termed 'data'. But need data to be observational? Could they not be mathematical or theoretical? And need they be about real-world phenomena?

Re-emphasizing the need for a new definition, Hjørland (2018, 690) points to the definition of data in the form of unit of analysis by quoting Kaase (2001): "Data is [are] information on properties of units of analysis." Here to be better termed "a unit of analysis" rather than a variable, but we cannot here open an analysis of the concept "variable". To overcome the discrepancy among different definition of data, Hjørland (694) suggests that social epistemology is the most fruitful approach for understanding data.

The diverse range of epistemological perceptions related to the concept of data, although indicative of differences in understandings and the definitions, at the same time expresses the intrinsic capability or talent, and consequently, the interpretative variety of the data concepts. As such, there seems to be a great deal of opportunity to define and theorize data from different perspectives. To formulate theories about data one must, first of all, recognize and understand the different features and capabilities different human beings or human-made machines can derive from data.

To come to a common definition relevant to the aim of this paper, it can be said that the concept of data has a spectrum of meanings ranging from an abstract element to a real-world fact having a variety of potential capabilities. Where there are differences in perceptions and mindsets data can be perceived as an abstract concept, meaning that we, as human beings, assign meaning to data based on our subjective mental schemas and the context in which we are thinking. On the other hand, depending on the real-world settings, data cover objective elements being captured, processed and stored by living and/or non-living systems such as humans/animals or machines. Here in this paper we perceive data as a captured or taken element not a given one. Once such an element is understood and recorded in our minds and/or in computerized systems to create a connection, then the data becomes objective and a reality understandable by others (we call this a 'fact', such as units, digits, symbols which can be analyzed and processed). This concept about data, that is, data "not as a given but as a taken" element shows its real feature in our real world. Hjørland (2018, 686; quoting Machlup 1984 and Drucker 2011) emphasizes the similar notion as a significant epistemological issue.

5.0 Features and capabilities that can be derived from data

Before explaining any features and capabilities, it must be emphasized again that all data capabilities are dependent on human physical and mental processing including activities of human-made machines. This, as will be discussed later in detail, implies that data by itself has no meaning or importance unless human beings give to or get the desired meaning and/or usage from data.

That said, the following list of data capabilities can be considered as objective contexts for data theorizing:

1. Data can be "captured and collected". Data is an element to capture. It can be extracted from various sources of

production (possibly any entities and events in the real world) in various ways and stored for later processing.
2. Data can be "saved". This is one of the most obvious and useful features of data as a taken element. This indicates that, something which can be stored (in any format or on any medium), can potentially be useful for later processing and subsequent consumption.
3. Data can be "processed". Processing is the basis for human beings and organizations' activities such as analysis, reasoning, and inference (data mining and knowledge extraction). Different types of processing, from simple to complex, can be performed on the data. By processing the data we can create information and/or knowledge.
4. Data can be "transferred and shared". This feature enhances the value of data in various inter-personal and inter-organizational communications because it enables the transfer and sharing of data between and among individuals or organizations or systems, regardless of time and place limits. Data sharing is now a common trend for many government, research or commercial organizations.
5. Data can be "bought and sold". This feature indicates the importance and value of data as a commodity. From this perspective, data can even be a strategic commodity for some organizations especially governments and big companies.
6. Data can be "used and re-used". One distinctive feature of data is that it can be used repeatedly. Data produced or generated by any individual, organization or tool can repeatedly be used for similar or different purposes, even by others. This is an indication of the fact that data has huge added values.
7. Data can be "measured". This ability is especially important for professional and scientific organizations and researchers. By using different tools and different methods data can be measured in terms of the quantity and quality of production, processing, transmission, composition, and reproduction. Also, from another point of view, we can measure data in terms of accuracy, validity, up-to-dateness, and the like. Data as metrics can be a significant approach to measure the level and amount of data activities (such as capturing/collection, processing, organization, transfer, and use) in organizations.
8. Data can be "refined". It is obvious that not all the produced or captured data are of the intended quality for individuals and organizations. For this reason, there is a tendency or obligation for individuals and organizations to refine and filter their data based on their policies and priorities.
9. Data can be "dynamic". This feature is part of the nature of the data. Contrary to the original notion (i.e., data defined as a raw material, as stated earlier), data can be a dynamic (active) element and can be changed from its static state to produce new data or be manipulated and combined with other data.
10. Data can be "communicated and understood". Man's mental ability to process data has given him the ability to communicate with others and create meaning. This is achievable through linking/integrating the captured data to the information and knowledge already stored in his mind.

As a general conclusion regarding the different aforementioned capabilities and features, data is so valuable and important that almost all individual, managers, experts, and researchers in public and private organizations are heavily dependent on it; without access to data research can hardly be carried on and the survival of many organizations is under serious threat. So the significance of the values and the different capabilities of data leads us to the necessity of developing theories about it.

6.0 Why do we need theories about data?

There are some criticisms regarding the lack of attention to theory and theory building in the LIS discipline. Pierce Butler was among the first who stated in his book *Introduction to Library Science* (1933) a lack of interest in librarians about theories in their profession. Buckland (1988) reaffirmed this in his notable book *Library Services in Theory and Context* and believed that there was barely any discussion of whether or not there might have been any progress in the development of theoretical understanding of librarianship since 1933.

Similarly, Hjørland (2015) states that the notion "theory" is a neglected concept in the field of information science (IS) and knowledge organization (KO) as well as generally in philosophy and in many other fields, although there are exceptions from this general neglect (e.g., the so-called "theory theory" in cognitive psychology).[3]

The important point regarding the need for developing theories for data is that, while there are some theories about the concept of information and knowledge almost beyond the field of IS, little effort has so far been made to develop general, scientific, and/or philosophical theories about data. We have not addressed the significance of data in the context of scientific theories based on inviolable laws or principles. Capurro and Hjørland (2003) believe that raising the subject of data has been much delayed in the field of Information Science. Borgman (2007) and Furner (2016), too, state that the topic of data has been neglected in many Information Science texts. Furner (2016) points out that the place for a data review article, like the paper published by Capurro and Hjørland (2003) is missing. One possible reason may be that the different diverse definitions set forward by different scholars on one hand and the ambiguous nature of data on the other hand, has added to its complexity. This in itself requires an in-depth look at this vital element from a theoretical point of view.

6.1 Theories for the data

Theories set systematic rules about the phenomena in the universe and the relationships between/among them. To formulate any theory about data, first of all one needs to consider the relevant frameworks, conditions, and requirements of theory development. To introduce an idea as a theory we must have the necessary frameworks for theorizing (such as systematic thinking and defining the relationship between the facts, concepts or elements). Merriam Webster's International Dictionary (2020) provides some of the essential requirements related to the ability of an idea to theorize which are implied in the definition of the "theory" itself:

1. a plausible or scientifically acceptable general principle or body of principles offered to explain phenomena;
2. the analysis of a set of facts in their relation to one another;
3. a hypothesis assumed for the sake of argument or investigation;
4. the general or abstract principles of a body of fact, a science, or an art.

It should be added that there are various views on the conditions for developing reliable theories in different fields. Deductive and inductive methods of analyzing, explaining and reasoning are basic requirements for theory building.

Given the diverse potential capabilities and the valuable role of data in the lives of individuals and organizations on the one hand, and the growing significance of data science on the other, it seems that the necessary ground is ready for developing theories about data. In general, the purpose of formulating theories is to gain a better and deeper understanding of the whats and whys as well as the considerable potential capabilities of data, especially metadata and big data, as a new and vital fact in the world. Developing relevant theories can enrich and enhance many data-centered and data-related fields, particularly data science, information science, knowledge science, cognitive sciences, computer sciences, life sciences, and management.

Based on the definitions of theory mentioned earlier, here are some general theoretical propositions, as a starting point, on which we can develop new theories about the data. These propositions provide logical explanations about the facts and phenomena as well as the association between or among phenomena in which data has the crucial role:

1. Similar to information and knowledge, data is not an independent or self-existent element. Data cannot be considered meaningful independent of its recipients. It should be captured and processed. Capturing, processing, understanding and using data is dependent on human beings or man-made systems (such as machines or computers). Likewise, understanding objects, events, and concepts is basically dependent on the intentional capturing and processing of data by the nervous system.
2. As long as human beings or human-made machines capture and transform the signals of their interactions with or observations of the environment into understandable elements, data are constantly produced, processed, consumed, and reproduced as facts. There is potentially no limit to the production and growth of data through events and human activities. Data is an ever-growing element which gradually generates large corpuses namely big data. We are facing the era of overwhelming growth of data, metadata and big data. From an epistemological point of view this overwhelming increase in the volume and types of data is forming a new horizon as well as new problems for human beings.
3. To human beings and human-made machines data are facts representing the quantity and/or quality of living and non-living objects and events. It is through the captured data that we are able to understand entities and how they work. As such, data act as evidence of entities. Thus, every being or every event has the potential to produce a sign (potentially, data) by which we are able to identify them. From this perspective, i.e., within the human ability to perceive, whether real or imagined, data is evidence to describe and represent the quantity and quality of the real world and ecosystems. It is through data that humans or human-made machines are able to recognize and record physical/numerical variables, such as seismic data from earthquakes and volcanoes, or wavelengths of light coming from stars and distant galaxies, types and numbers of elements in a material, blood test results, etc.
4. The formation and understanding of many types of relationship between or among entities in the real world is mainly based on the ways and techniques that data are captured, processed and analyzed. In other words, many types of relationships can only be formed and understood through taking the role of data into consideration, that is, through capturing, processing, and analyzing data. All the real-world objects and all ecosystems exist based on an entity-relationship pattern. This is a significant concept on which we are able to discover and understand the basic requirement of our existence in the ecosystems around us.

5. Data, information, and knowledge are inter-related and interdependent elements, so they have interactive behaviors. They can be transformed into one another in cases where humans or human-made machines decide. Also, much of the data stored in IR systems is the result of information and knowledge that human beings have produced through their activities. Big data is an example. Information and knowledge must be de-structured, re-formatted, recorded, and stored in the form of data (in database fields) so that they can be indexed, saved, transferred, retrieved and used. The more objective and sophisticated data processing, the more useful information and knowledge will be produced.
6. Based on the above proposition, the discovery and understanding of the relationship between data, information, and knowledge is basically dependent on human beings or human-made systems. This relationship between or among the three concepts is primarily intellectual (i.e., it is formed in the minds of humans) or is based on inference computer programs. It is through human reasoning system or data collection and processing by human-made machines that data becomes meaningful and usable. Hence, its reliability and durability are to be received, understood, and stored on embodiments such as the human brain or human-made tools.
7. Human behaviour in the production, processing, and usage of data, especially at the organizational and social levels, and in the case of big data, are dynamic, meaningful, and savable phenomena. These behaviors can be studied epistemologically, philosophically and scientifically as phenomena. As noted earlier, much of the activities in today's society is based on data capturing and processing. Data-driven approaches are the dominant trends for many individuals as well as many institutions and companies. Big data and linked data are, in fact, evidence for such trends at the large scale. The field of study of such trends, using data, may be called "sociology of data". In this regard, it is important to identify how the human groups and social institutions (whether educational, research, industrial, commercial, and so on) capture, process, transfer, share, and consume data. We can also identify the nature of any phenomena and the causal relationships between them from different dimensions of the data.
8. The level of activities of living organisms (such as human beings and organizations) normally depends on the quantity and quality of data capturing, processing, synthesizing, and interpretation by them. The more active and sophisticated an organism, the more it is able to capture, process, produces, and transfer data. This means that there is a relationship between the level of data processing by living systems and the level of data utilization for producing information and knowledge. In fact, the degree of activeness (or dynamism) of an entity depends on its ability to capture and process data. For example, in comparison to many other animals, a human being is a creature who has a more dynamic and complex mind able to generate data, information and knowledge. For this reason, human beings have more interactions and higher mental activities that result in producing more and more diverse data needed for decision making and action. The same argument can be made for large and important organizations compared to smaller ones (for example, aerospace and aviation organizations, or large and reputable universities and research laboratories in comparison to a small office or a high school). Likewise, more complex and sophisticated devices, such as aircrafts compared to cars, and cars as compared to bicycles, require more data for their building and performance. Based on this argument, it can be said that by analyzing and measuring the extent and types of data produced and processed by each entity, one can measure the level of its activeness and complexity (data as a measurement unit or variable). Also, more complex and dynamic entities are able to better understand the relationships between concepts and events based on the processed data. So it can be hypothesized that "the more active organizations and individuals, the more data they capture, process, consume, produce and/or reproduce. Likewise, the more data is produced, the greater is its impact on the activity, structure and processes of the organization.

In summary, in connection with the dynamism of the data we can theorize that all human beings and organizations potentially can capture, process and produce data, but the level of activeness of any entity depends on the extent and sophistication of the mental and/or computational processing. Here data can be regarded as metrics.

7.0 Conclusions

Discovering and understanding the real world, its structure, and its constituents (entities, events, as well as the relationships among them) requires an understanding of the nature and philosophy of data. Data are the essence and basic elements for interaction with the world, understanding things and for decision making in life. If we accept the notion that data is the evidence of how we understand and mentally process our observation of and interaction with the environment, then it can be argued that all the entities and events can potentially produce signals (data). Since all living systems are in interaction with the environment (i.e., other entities and events), there is no limit to the production and growth of data. As long as entities and events exist in the world, data is produced, consumed, and reproduced to give meaning to life and events. As mentioned earlier, as any entity which is more dynamic and more sophisticated, it can generate, process and transfer more data. It seems that with the advent of new technologies, the dynamism of human beings and organizations will increase and this, too, will produce more data. Dynamism means more activities and more processing capabilities by an entity. Based on this argument, it can be said that by analyzing and measuring the quantity and quality of data produced by each entity, we can measure its dynamism and sophistication. Big data and linked data are evidences of this dynamism at the macro level.

It is crucial to understand that data, information, and knowledge are associated and interdependent concepts. Any theoretical and philosophical inquiry into any of these three concepts should take such association into consideration. As noted earlier, there are some theories developed so far about information and knowledge but not about data. The present era is witnessing the provision of suitable contexts for the development of the theoretical and practical concept of data (data science and all related fields). From a scientific point of view, data is the most important empirical evidence in the activities of scientists and the development of science and knowledge. In other words, the development of scientific theories in many fields is dependent on the collection, processing, analysis, synthesis, and inference of data.

This paper presented a number of general propositions for theories as a starter list with many other aspects yet to be explored about data. These theories, in turn, can be considered as meta-theories to formulate specialized theories of data in specialized domains. The expansion of these theories can provide a good framework for the theoretical and practical development of data science. The present paper attempted to argue that understanding the universe, entities, and events, as well as the relationships between them, requires an understanding of the theoretical and practical basis of the data. Such an understanding can further enhance data science, create new educational and research trends, more accurate findings on the relationships between entities and phenomena, and ultimately, a deeper understanding of the world.

In summary, we will see major developments in the realm of data and data science, and this requires deep theoretical and applied studies. Human society is now on the verge of a valuable, universal data paradigm and revolution, a development that has opened new horizons in all areas of human activities. Research and practice areas appear to be emerging and, consequently, new specialties based on data capabilities. It can even be expected that something new, such as data ethics, will become an important issue in human societies because, as with information, many challenges such as misunderstanding of data, misuse of data, and the like, can be harmful. In the meantime, theorizing and formulating philosophical and scientific theories related to data have a special place. The important notion supporting the development of theories about data is that, if data science is to continue flourishing in a variety of specialized fields and trends, then we need to build theories about data for research and practical purposes in a multi-disciplinary context.

Notes

1. An entity can be any existence (living or non-living) separate from others. Examples are a human being, an animal, an organization, a building, a tree, a book, or even a concept, etc. According to the Merriam Webster Dictionary, entity is something that has separate and distinct existence and objective or conceptual reality. Here in the context of this paper, entity refers to living existences only.
2. All the happenings in the environment and all activities done by humans are examples of events. According to Merriam Webster Dictionary, an event is something that happens, a postulated outcome, condition, or eventuality, such as a social occasion or activity.
3. Theory-theory (or theory theory) is a scientific theory relating to the human development of understanding about the outside world. This theory asserts that individuals hold a basic or naïve theory of psychology (folk psychology) to infer the mental states of others (Ratcliffe 2006 quoted in Wikipedia “Theory theory”).

References

Abiteboul, Serge. 1997. “Querying Semi-Structured Data.” In *Database Theory - ICDT '97*, edited by F. Afrati and P. Kolaitis. Lecture Notes in Computer Science, vol 1186. Berlin, Heidelberg: Springer, 1-18. https://doi.org/10.1007/3-540-62222-5_33

Ackoff, Russell Lincoln. 1989. “From Data to Wisdom.” *Journal of Applied Systems Analysis* 16: 3-9.

Akude, John E. and Philip Grunewald. 2014. “Knowledge for Development: Findings of a Stakeholder Consultation Facilitated by the German Development Institute (GDI/DIE).” *Knowledge Management for Development Journal* 10, no. 3: 113-30.

Atkinson, R.L., R.C. Atkinson, E.E. Smith, D.J. Bem and S. Nolen-Hoeksema, S. 2000. *Hilgard's Introduction to Psychology* 13th ed. New York: Harcourt College Publishers.

Ballsun-Stanton B. and Debora Bunker. 2009. “Philosophy of Data (PoD) and its Importance to the Discipline of Information Systems.” In *AMCIS 2009 Proceedings* 435: 1-8. San Francisco: AIS Electronic Library (AISeL), 1-8. http://aisel.aisnet.org/ amcis2009/435

Baars, B. and N. M. Gage. 2013. *Fundamentals of Cognitive Neuroscience: A Beginner's Guide*. London: Elsevier.

Berlinger, Gene, Durval Castro and Anthony Mills. 2004. "Data, Information, Knowledge and Wisdom." http://www.systems-thinking.org/dikw/ dikw.htm

Borgman, Christine L., Jillian C. Wallis and Noel Enyedy. 2007. "Little Science Confronts the Data Deluge: Habitat Ecology, Embedded Sensor Networks, and Digital Libraries." *International Journal of Digital Library* 7: 17–30. https://doi.org/10.1007/s00799-007-0022-9

Borgman, Christine L. 2010. "Research Data: Who Will Share What, With Whom, When, and Why?" Unpublished paper, presented at China-North America Library Conference, Beijing 2010. https://works.bepress.com/borgman/238/

Brooks, David. 2013. The Philosophy of Data. *The New York Times* February 4. http://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html

Buckland, Michael K. 1988. *Library Services in Theory and Context*. New York: Pergamon Press.

Butler, Pierce. 1933. *Introduction to Library Science.* Chicago: Chicago University Press . [Reprinted in Richardson (1992)].

Cao, Longbing. 2017. "Data Science: A Comprehensive Overview." *ACM Computing Surveys* 50(3): 1-42. https://doi.org/10.1145/3076253

Capurro, Rafael and Birger Hjørland. 2003. "The Concept of Information." *Annual Review of Information Science and Technology* 37: 343-411.

Capurro, Rafael. 2009. "Past, Present, and Future of the Concept of Information." *tripleC: Communication, Capitalism & Critique* 7: 125-141.

Carnahan, Karlyn. 2014. "The Importance of Data Governance: Current Practices." http://celent.com/reports/importance-data-governance-current-practices

Choenni, Sunil, Niels Netten, Mortaza S. Bargh and Rochelle Choenni, 2018. "Challenges of Big Data from a Philosophical Perspective." In *International Conference on Multidisciplinary Research (MyRes), Mauritius,* 71-8. https://doi.org/10.26803/MyRes.18.06.22

Choo, Chun Wei. 1996. "The Knowing Organization: How Organizations Use Information to Construct Meaning, Create Knowledge and Make Decisions." *International Journal of Information Management* 16: 329-40.

Dessaei, S. F., N. M. Rane, B. S. Al-Shammari, S. H. Al-Sabea and M. Al-Naqi. 2019. "Challenges in Operating a Digital Oilfield - Lessons Learned From The Burgan Integrated Digital Field Pilot," Paper presented at the SPE Kuwait Oil & Gas Show and Conference, Mishref, Kuwait, October 2019. Paper Number: SPE-198029-MS. Society of Petroleum Engineers. doi:10.2118/198029-MS

Dewey, John. 2019. *The Logic of Human Mind, Self-Awareness and the Way We Think*. Musaicum Books. London: Madison & Adams.

Donoho, David. 2017. "50 Years of Data Science." *Journal of Computational and Graphical Statistics* 26:745-66.

Drucker, Johanna. 2011. "Humanities Approaches to Graphical Display". *Digital Humanities Quarterly* 5, no. 1: (electronic source, no pages; 52 sections). http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html

Fattahi, Rahmatollah and Ebrahim Afshar. 2006. "Added Value of Information and Information Systems." *Library Review* 55, no. 1/2: 132-47.

Floridi, Luciano. 2008. "Data." In *International Encyclopedia of the Social Sciences. 2nd edition,* vol. 2, edited by W. A. Darity. Detroit: Macmillan, 234-7. http://www.philosophyofinformation.net

Floridi, Luciano. 2012. "Big Data and Their Epistemological Challenge." *Philosophy and Technology* 25:435–7.

Franks E. W. 1966. "A Data Management System for Time-Shared File Processing Using a Cross-Index File and Self-Defining Entries." In *AFIPS '66: Proceedings of the Spring Joint Computer Conference.* New York: ACM, 79-86.

Franksen, O.I. 1985. "The Nature of Data - From Measurements to Systems." *BIT Numerical Mathematics* 25: 24-50. doi:10.1007/ BF01934986

Furner, Jonatan. 2016. "Data: the Data." In *Information Cultures in the Digital Age: A Festschrift in Honor of Rafael CapurroI*, edited by M. Kelly and J. Bielby. Wiesbaden: Springer, 287-306. doi:10.1007/978-3-658-14681-8_17

Furner, Jonatan. 2017. "Philosophy of Data: Why?" *Education for Information* 33: 55–70. DOI 10.3233/EFI-170986

Gil, Press, 2013. "A Very Short History of Data Science." *Forbes Tech Blog* (last updated May 28, 2013). http://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/2/

Hakeem, B. 2016. "The Business Value of Data." *Sealed Air Blog*. https://sealedair.com/blog/ business-value-data.

Hjørland, Birger. 2015. "Theories are Knowledge Organizing Systems (KOS)." *Knowledge Organization* 42: 113-28.

Hjørland, Birger. 2018. "Data (with Big Data and Database Semantics." *Knowledge Organization* 45: 685-708. DOI: 10.5771/0943-7444-2018-8-685

Horne, Jeremy. 2018. "Visualizing Big Data from a Philosophical Perspective." In *Handbook of Research on Big Data Storage and Visualization Techniques*, edited by Richard S. Segall and Jeffrey S. Cook. Hershey, PA: IGI Global, 809-52. https://doi.org/10.4018/978-1-5225-3142-5.ch028

Jifa, Gu and Zhang Lingling. 2014. "Data, DIKW, Big Data and Data Science." *Procedia Computer Science* 31: 814-21.

Kaase, Max. 2001. "Databases, Core: Political Science and Political Behavior". In *International Encyclopedia of the Social and Behavioral Sciences*, edited by Neil J. Smelser and Paul B. Baltes. Amsterdam: Elsevier, Vol. 5, 3251–5.

Labrinidis, Alexandros and H.V. Jagadish. 2012. "Challenges and Opportunities with Big Data." *Proceedings of the VLDB Endowment* 5: 2032-3. 10.14778/2367502.2367572.

Leonelli Sabina. 2016. "The Philosophy of Data." In *The Routledge Handbook of Philosophy of Information*, edited by L. Floridi. London: Routledge, 191-202.

Machlup, Fritz. 1984. "Semantic Quirks in Studies of Information". In *The Study of Information: Interdisciplinary Messages*, edited by Fritz Machlup and Una Mansfield. New York: Wiley, 641–71.

Maycotte, Higinio O. 2015. "Quantifying the Value of Effective Data." *Forbes* (January 27, 2015). https://www.forbes.com/sites/homaycotte/2015/01/27/quantifying-the-value-of-effective-data/#54f26b847872

Merriam-Webster. (n.d.). "Theory." In *Merriam-Webster.com dictionary*. https://www.merriam-webster.com/dictionary/theory

Mons, B. et al. 2011. "The Value of Data." *Nature Genetics* 43**:** 281–3. https://doi.org/10.1038/ng0411-281

Nielsen, Hans Jørn and Birger Hjørland. 2014. "Curating Research Data: The Potential Roles of Libraries and Information Professionals." *Journal of Documentation* 70: 221-40. 10.1108/JD-03-2013-0034

Pouchard, Line, Amy Barton and Lisa Zilinski. 2015. "Data Narratives: Increasing Scholarly Value." *Proceedings of the American Society for Information Science and Technology* 51, no. 1: 1-4. https://doi.org/10.1002/meet.2014.14505101088

Ratcliffe, Matthew. 2006. "Folk Psychology is Not Folk Psychology." *Phenomenology and the Cognitive Sciences 5, no. 1: 31–52.*

Rowley, Jennifer E. 2007. "The Wisdom Hierarchy: Representations of the DIKW hierarchy", *Journal of Information Science* 33: 163-80.

Shahzad, Wassi, Qamar Rehman and Ejaz Ahmed. 2017. "Missing Data Imputation Using Genetic Algorithm for Supervised Learning." *International Journal of Advanced Computer Science and Applications* 8: 438-45.

Shwetha, T. M. and B. L. Muralidhara. 2018. "e-Governance for India's Development: A Review." *3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India*, 214-20.

Sun, Zhaohao and Kenneth Strang. 2018. "The Philosophy of Big Data." PNG UoT BAIS 3(8): 1-8, DOI:10.13140/RG.2.2.16139.80161

Swan, Melanie. 2015. "Philosophy of Big Data: Expanding the Human-Data Relation with Big Data Science Services." *2015 IEEE First International Conference on Big Data Computing Service and Applications* Redwood City, CA: IEEE, 468-77. doi:10.1109/BigDataService.2015.29.

Torra, Vicenç. 2003. "Trends in Information Fusion in Data Mining." In *Information Fusion In Data Mining*, edited by V. Torra. Berlin; Heidelberg: Springer, 1-6.

Undaiva, Jaimin, Atul Patel and Nikhil P. Shah. 2015. "Education Data Mining - A Prediction Model and Its Affecting Parameters for Prediction in Higher Education," *International Journal of Engineering Research & Technology (IJERT)* 7:1-5. DOI:10.13140/RG.2.1.4514.1840

Van der Aalst, W. 2016. "Data Science in Action." In *Process Mining*. Berlin: Springer, 3-23 doi.org/10.1007/978-3-662-49851-4_1

Vandeventer, R. 2016. "How to Determine the Business Value of Data." http://dataclairvoyance.com/ blog/how-to-determine-the-business-value-of-data

Vertesi, Janet and Paul Dourish. 2011. "The Value of Data: Considering the Context of Production." *Proceedings of the ACM Conference on Computer Supported Cooperative Work,* CSCW, 533-42.

Weinberger, David. 2010. "The Problem with the Data-Information-Knowledge-Wisdom Hierarchy." *Harvard Business Review*, February 02. https://hbr.org/ 2010/02/ data-is-to-info-as-info-is-not

Wiener, Norbert. 1950. *The Human Use of Human Beings.* New York: Houghton Mifflin.