Modelling the Universe of Data (DataVerse)

Ashutosh Gupta

GM, Gujarat Telecom Circle, BSNL, Ahmedabad.

Abstract- The paper presents an axiomatic model of Abstract Universe or the Universe of Information. It suggests the minimum elements required to create the information of Abstract Universe and this model forms the necessary and sufficient condition for an Abstract Universe to exist. This Universe of Information is viewed for outside to see its impact on real Universe. The process of creation of Information Universe from its Axiomatic model is proposed. Some interpretation based on the model is proposed. This model will form the basis for the axiomatic model of the Real Universe which is beyond the scope of the paper.

Index Terms- DataVerse, Data, Process, Algorithm, Datafication, Instantiation, Actor, Agent, Axiomatic Model, RootCause

I. INTRODUCTION

In today's information age, information is multiplying exponentially and the devices using these data are exploding. Various data sources as Internet of Things (IoT), sensors, user devices as mobiles etc. are creating enormous data. Analysis of this data as Big Data, Artificial Intelligence (AI) and Deep Learning are the attempts to map the reality and to emulate it. This growing is creating an increasing Virtual Universe and Abstract Universe. The advent of communication and Information Technology also has exploded the Data Universe. With such an impact of Data in our lives, it has become important to understand tha nature of the Data Universe in a holistic way.

II. WHAT IS DATA

Data comes from the Latin term meaning literally 'something given,'. It is a set of codes that represent information, processes, measurements, any phenomenon, any physical or non-physical aspect, etc. As such, data as a general concept refers to the fact that some existing information or knowledge is represented or coded in some form suitable for better usage or processing. Data is always a slice of reality, we are chopping the Universe up into little bits, codifying into symbols and storing them to be utilised in converting it back to Reality. Such data can be brought back to reality.

This makes the data portable. It can be taken from one context and brought into another. No matter how extensive our gathering of data it will only ever be able to capture a partial representation of the system and in so doing separate it from its overall context. Data is always incomplete but the discrete nature of data makes it quantifiable and thus accessible to formal quantitative methods of analysis (Colchester, Datafication -

Complexity Labs, 2018). *The Universal Set of all such data can be termed as* **DataVerse.**

III. WHAT CONSTITUTES THE DATAVERSE

Any information which is captured, recorded and stored will constitute the DataVerse. This data can be information (meaningful) or noise (meaningless). These information can be any one of the following.

1. Literature, Language, Books, Documents, etc. which are presented in the form of documents or books etc.
2. Sound or audio stored in Cassettes, Memory devices, tapes etc, which can be played back.
3. All types of pictures, photographs
4. Movies, Videos, etc.
5. Data on the internet, Social Media, etc.
6. Memory in the mind of living being which can be recollected.
7. Any operator or Process or Algorithm which has been designed for the processing, modification, manipulation or replication or the data.
8. All the data and operator for its processing constitutes knowledge also forms the part of DataVerse.

IV. CREATION OF DATA

The data can be created in multiple ways. Some of them are

1. Act of Measurement or Datafication: It is process of taking snapshot of Reality, codifying it and storing it. The snapshot can be of physical objects, process, phenomenon, ideas, thoughts, feelings, etc. Any aspects of reality can be converted to Data by the process of Datafication. (Colchester, 2018). E.g. Recording a movie is an act of datafication of the event.
2. Manipulation of existing Data/Applying Operator on the Data: The existing data can undergo various operations (as arithmetic, logical, process, algorithm, etc.) to create more data. The processing device can be human mind or a computer. Generally such processing creates the data of higher order. Eg Graph from raw data, words from alphabets, etc. A language is a recursive process of data manipulation on alphabets. E.g. Giving the VFX to the movie is the manipulation of movie.
3. Replication of Data: A data can be increased by replicating or copying. A broadcast is one of the example of replicating in large quantity. In today's world, data on

internet, Youtube, Facebook, Newspaper etc. are big replicators. Because of such replications, we are overwhelmed with data in our day to day life. E.g. Broadcasting or Downloading the movie is the process of replication.

Destruction of Data

The data which is created can be destroyed. It can be destroyed in the following ways

1. Destruction of storage device where data is stored: If the storage device is destroyed, then the data stored in it is destroyed.

2. Destruction of the Interpreter or Instantiator (Cohchester, 2016): Since the data are set of symbols, it needs interpreter to make sense. If the interpreter of data is absent or destroyed, the data cease to exist. For example, the data in a cassette cease to exist if there is no player. A language is destroyed if there is nobody to read or understand it.

Axiomatic Model of DataVerse

Since the data is the Symbolic or the coded version of the real Universe, it is possible to reduce all type of data into its Elementary Code by breaking the data into simpler form till the simplest indestructible Elementary Code is arrived. Following is some of the effort to reduce the data to its Elementary Code.

1. Literature → Book → Chapters → Paragraphs → Sentences → Words → Alphabets
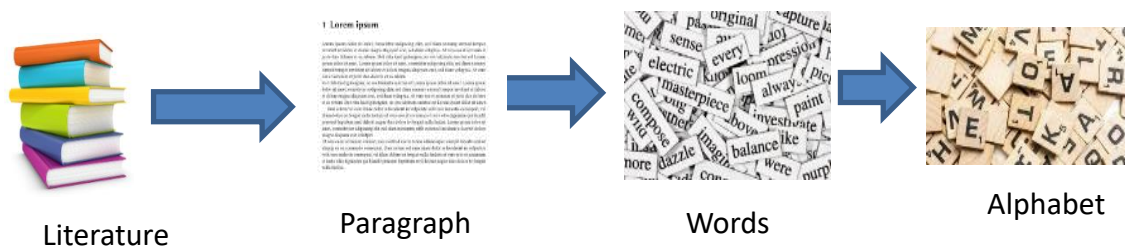

Figure 1   Reduction of Literature into Elements

The Elementary code derived are the alphabets which cannot be subdivided.

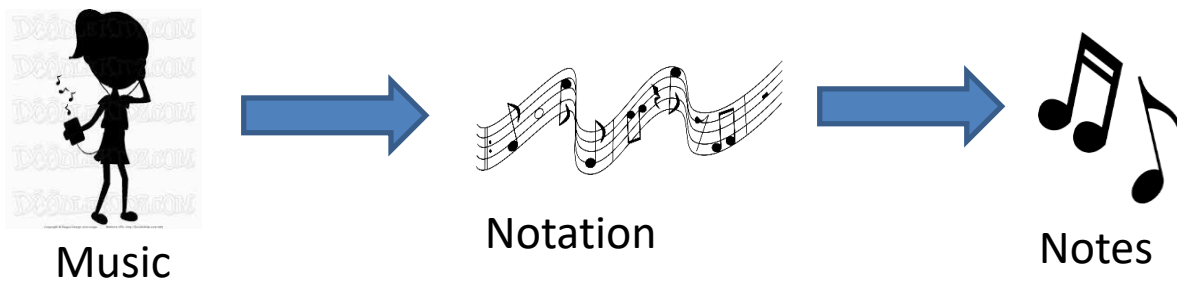2. Music → Notations → Note . The indivisible element is notes.

Figure 2   Reductio of Music into Elements

3. Movie → Pictures → Pallete. Here the element is the color palate which cannot be divided.

From the above examples it can be seen that for different type of information there were different indivisible elements. Hence there is the need of a model which can represent all types of elements.
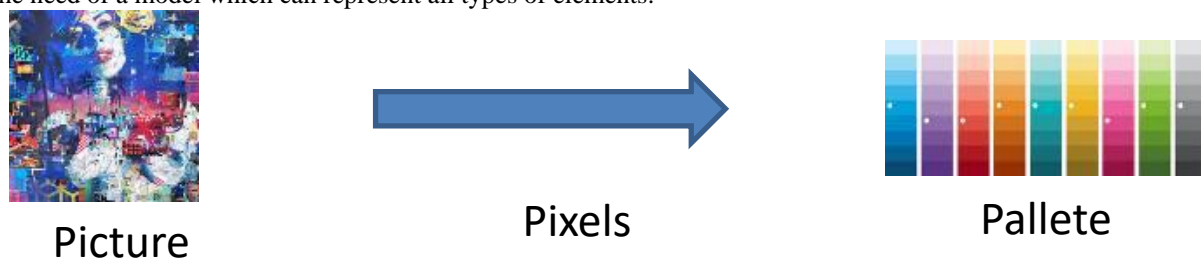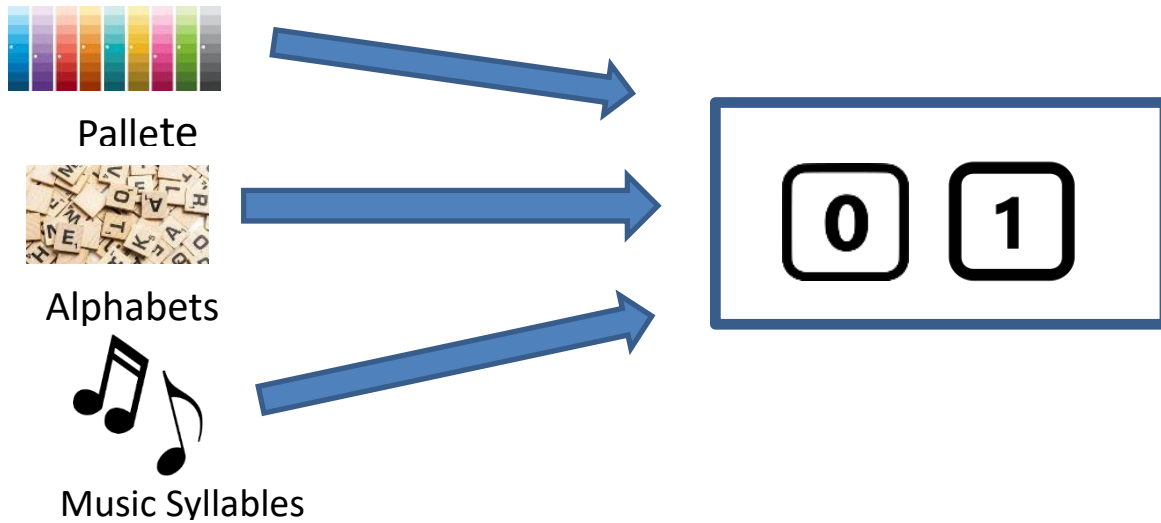

Figure 3   Reduction of Picture into Elements

With the advent of Computers and Binary representation of Information, all the information can now be represented in terms of Bits('0's and '1's). Now all the forms of Information as Alphabets, Pallets, Music notes can now be represented in the combinations of Bits.

Bits(**0,1**) are the Elementary Code and all the data in this DataVerse which is the interplay of these basic Elementary Codes. These

Figure 4   Representation of various Infromation in the terms of Bits(0,1)

Elementary Codes are perfect Source and Sink (i.e. infinite numbers of 0's and 1's can be extracted from these Elementary Codes and all the symbols can be absorbed into it).
From the above reductionist approach, the Elementary Codes of the DataVerse can be
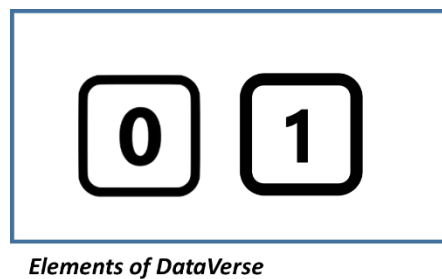represented as

Elements of DataVerse

Figure 5   Elementary Codes of DataVerse

The DataVerse is composed of two Elementary Codes Bits(**0,1**). All the data in the DataVerse is the interplay of these Bits(**0,1**). All data cease to exist once these symbols are in equilibrium state(0 remains 0 and 1 remains 1). This Equilibrium state of the DataVerse is called **Unmanifested form of DataVerse or the RootCause of the DataVerse.** It has the potential of becoming Universe of Data or DataVerse.

This DataVerse is static in nature. The data inside the DataVerse cannot modify itself. In order to have dynamism in the data, there is need of another property (Motion) to make any modification in the data effective. The DataVerse is dynamic only to the extent, the motion acts upon the DataVerse. It is just like a vehicle without fuel. It moves to the extent it is pushed. It doesn't move by itself. This Motion Element provides the dynamism to the entire DataVerse. For e.g. a child picks up the words and forms a sentence. The words were static before and now the sentence is static and both are data. There was small component of motion by the child which changed the state of the DataVerse. Hence there is a new Element Motion which when acts upon DataVerse changes the state of the DataVerse. As the data in the DataVerse do not change by itself, this Motion Element is external to the DataVerse as shown below
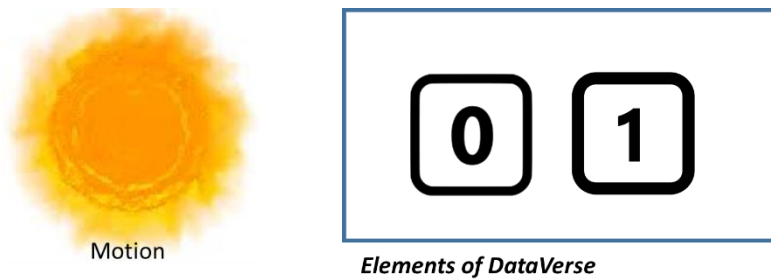
Figure 6: DataVerse when Motion acts on it

The above scenario also does not make the sense unless there is an Actor / Agent supporting the motion to make sense to the data.

Actor/Agent – An Actor is an entity responsible for all the action. The following are the properties of an Agent

     a.   It is the representative of I or Ego.
     b.   It can set Goals.
     c.   It can analyse the information, compare them, have options and make choices.
     d.   It is the THE FREE WILL.
     e.   It is responsible for the ACTION.

In the above example, the child's action enables the formation of the sensible sentences from the words. Hence the new Model of DataVerse becomes as below.
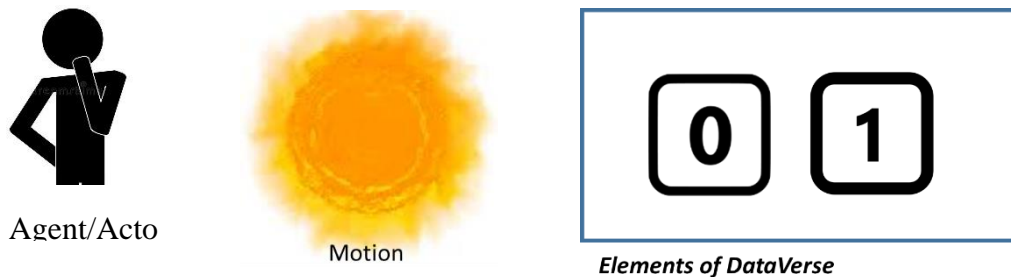
Figure 7: Axiomatic Model of DataVerse

The above scenario makes the DataVerse to be meaningful. Whenever there is an Actor acting on the DataVerse, the DataVerse changes State. Actor/Agent and Motion are external to the DataVerse.

Thus with reference to Figure 7, DataVerse is composed of the two Elementary Codes. The Motion element which is external to the DataVerse provides the dynamism and changes the state of the DataVerse. The Agent which is also external to the DataVerse is the Purpose of DataVerse. The entire DataVerse happens for the purpose of the Agent.

Thus the DataVerse can be defined as an *Interplay of Two elementary Codes acted upon by Motion for the purpose of Agent.*
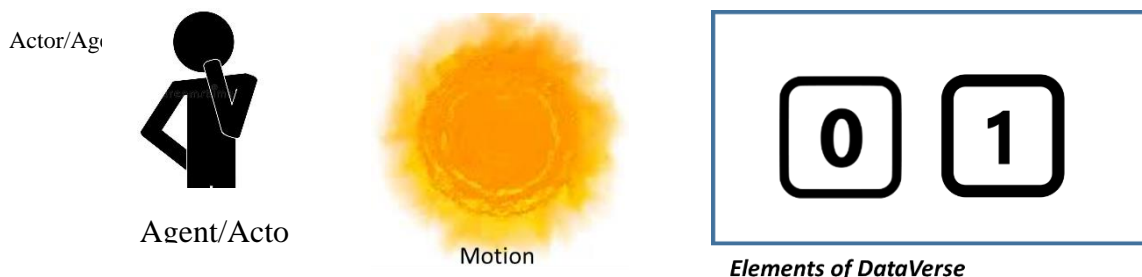

Figure 8 Model of DataVerse

Necessary and Sufficient Conditions for the existence of DataVerse

The above Axiomatic model of the DataVerse is both necessary and sufficient for the Creation of the DataVerse. The DataVerse will cease to exist in the absence of any one of the component

1. In the absence of any one the symbol, the data won't exist. There is no data with only 1's on only 0's.
2. In the absence of motion element, no new combination of data will be formed. No change, no evolution, no creation, no destruction of data will happen.
3. In the absence of Agent, the data created will be random or noise without any meaning.
4. The above model can be understood from the following example. The minimum items require to make a painting are two colours (Two symbols), one brush(Motion element) and a painter(Agent). The painting will cease to exist in the absence of any one of the element.

Hence the four elements are necessary and sufficient for the existence of the DataVerse.

Thus the model with its four elements is both necessary and sufficient for the existence of the entire DataVerse.

Properties of DataVerse

1. **Composition**: The DataVerse is composed of two Elementary Codes Bits(**0,1**). All the data in the DataVerse is the interplay of these Bits(**0,1**). All data cease to exist once these symbols are in equilibrium state.

Properties of Bits(0,1)

(a) The Bits(**0,1**) are omnipresent in the entire DataVerse.
(b) The Elementary Codes are Perfect Source and Perfect Sink.

2. **Static in Nature:** This DataVerse is static in nature. The data inside the DataVerse cannot modify itself since the Motion Element is not the component of DataVerse.
3. **Deterministic:** Since the DataVerse is Static in nature and the data in the DataVerse changes only when external motion acts upon it. The operation acting on the data called Algorithm is also a static data. Any algorithm acting on the data will give deterministic result. No randomness can be created in DataVerse.
4. **Discrete:** Since there are only two Elementary Symbols Bits(0,1) and no fractional symbols are created, the DataVerse is discreet and countable in nature.
5. **Finite and Countable:** Any subset of the DataVerse is finite and Countable.
6. **Unintelligent:** Since Agent/Actor is not the component of the DataVerse, DataVerse do not make any sense to its content.
7. Dependent on Unmanifested DataVerse for Creation or Absorption of Data.
8. All knowledge (set of Data and the Operators Action on Data) exist in the DataVerse.

Process of creation of DataVerse from the Axiomatic Model

The creation of the DataVerse happens by the process of Evolution and Emergence. More and more complex data is created by the successive evolution and emergence.

a) Creation of Language

The Elementary code inside the DataVerse in presence of Agent and Motion Element moves to produce patterns. The Agent chooses some of the patterns that make sense to form alphabets. Making sense is the property of Agent. The alphabets becomes the first level of Emergence for a language. The Agent in association with Motion Element Continues its interplay to make Words i.e. the combinations of alphabets that make sense, the Second level of Emergence. Other combinations which are not relevant to the Agent also exist in the DataVerse. This process of Emergence or Evolution continues to make Sentences, Paragraphs, Chapters, Books, Subject etc. The data in the DataVerse changes level of Emergence both in terms of depth and Diversity. The complete universe of data is formed from the above process of Emergence and Evolution. The DataVerse formed from this Interplay of the Elementary Code is called **Manifested DataVerse**.

b. Conway's Game of Life (Agent based modelling)

The process of creation is also seen in the Conway's Game of Life (Gardner, 1970), an agent based model. The bits in the Square of the Game is the Data. All the data in the game is all combination of 0's and 1's. The bits in the game are created or destroyed as per the need of the game. This data is static unless the tick happens. This tick is the motion element that acts on the data that changes the state of the data. Hence this motion is external to the DataVerse of the Game. This motion can be single tick or continuous ticks. The Data in the Square will change state so long the tick happens. At every tick, the data gets modified as per the rule defined by the Programmer. Here the agent does not take the decision but follows the logic of the programmer. Any pattern form in the data is based on the logic defined by the programmer. Here the real agent is the programmer which decides the how the data will change in the Game. It is again the programmer which sees the pattern in the data. Making sense is the property of the Agent. Different observers may identify different patterns. Based on the patterns the programmer may decide to change the logic of the game thus resulting into more patterns. This iterative process gives both emergence and diversity resulting into complex creative process.

V. CONCLUSION

This paper is an attempt to present the Information Universe as an Entity in the holistic perspective and named as DataVerse. It gives the crisp definition its nature and properties. The process of creation through emergence and evolution into a complex Information Universe. This model is also the seed to model the Real Universe. DataVerse together with the Real Universe can define the complete Reality (real and Virtual).

REFERENCES

[1] Cohchester, J. (2016, 07 30). Instantiation. Retrieved from Complexity Labs: http://complexitylabs.io/instantiation/

[2]   Colchester, J. (2018, 02 08). Datafication - Complexity Labs. Retrieved from Complexity Labs: http://complexitylabs.io/datafication

[3]   Gardner, M. (1970, October). MATHEMATICAL GAMES:The fantastic combinations of John Conway's new solitaire game "life". Scientific American, pp. 120-123. Retrieved 11 30, 2018, from https://web.archive.org/web/20090603015231/http://ddi.cs.uni-potsdam.de/HyFISCH/Produzieren/lis_projekt/proj_gamelife/ConwayScientificAmerican.htm

AUTHORS

First Author – **Ashutosh Gupta** was born in Badaun UP India, in 1969. He received the B.Tech. degree in Electronics & Communications engineering from the Banaras Hindu University, Varanasi in 1991and the M.E (EcE) degrees in Electrical Communications Engineering from the Indian Institute of Science(IISc) Bangalore, India in 1993.
In 1993, he joined the Department of Telecom, Govt. of India as India Telecommunications Services. From 1996 to 2000, he was instructor in Advanced Level Telecom Training Center, Ghaziabad. From 2000 to 2009 he was head of IT in Gujarat Circle, BSNL Ahmedabad. He was in-charge of pricing of Mobile Services for 4 years at BSNL Corporate Office New Delhi since 2009. Presently he is head of Wireline Business and IT in Gujarat Telecom Circle BSNL Ahmedabad.

Apart from his profession life, his interests are in studying recent development in Science and Technology, Complex Systems. In contrast, his interests are in studying ancient Indian Philosophy of Samkhya Yoga, Patanjali Yoga, Bhagwat Gita etc.
Email: bsnl.ashutosh.gupta@gmail.com
Mobile: 9426001919