The Future of Reality

Anna Arenas

University of South Florida

Abstract

Advances in quantum mechanics have pointed out that the driving force for all physical events is not the newtonian principle of conservation of energy but, simply put, the increase of entropy. At all levels, from subatomic particles to the world we see and to the cosmos, the increase of entropy is responsible for the “flow” of events.

Recent research in the quantum field indicates that objects can be interpreted only as events. Objects are not independent entities but a confluence of events. However, it doesn’t end there. Shannon’s research provided a gateway for a new paradigm to interpret reality. Since information measures entropy, it can be used describe any event -and consequently any object-. Therefore, the flow of what we understood in Newtonian physics as time, it is actually information..

Information is the underlying force behind every aspect of the physical natural world, but also of individuals and society. The present paper intends to analyze how our understanding of reality has been affected by the amount of information that is at our reach, thanks to the development of computational devices which have uncovered unprecedented amounts of data on every human activity. Statistics, laws of physics, nature, psychology, and sociology are all confluencing into a new paradigm that is challenging even our concept of individuality. As a sort of numerical determinism we find ourselves under the optic of human behaviour predictability inside network societies. It is also included in this brief analysis the relevance of finding adequate ways to deal with information for archival and retrieval purposes. For that matter a short review of the power of metadata and its intrinsic aspects are mentioned. Issues like the importance of Artificial

Intelligence (AI) and contextual effects are discussed as well, since both aspects are crucial for the future development of information systems.

The Future of Reality

Information has changed our interpretation of the physical world. It has transformed the way we recognize, understand, and deal with reality. It has even questioned our own free will, our sense of individuality, and our uniqueness. We are learning to see the world and ourselves with completely different eyes and construct meaning on new ways.

Ours has been called the Information Era. In our daily lives we have immediate access to mountains of data, we carry mobile devices that possess more computing capabilities than mainframes did just few decades ago, we use phones that run applications which understand voice commands and control objects located miles away. However, the most important reason this is considered the era of information it is not the computing revolution *per se*. Not that it is not relevant, without the advent of modern computers we couldn't have been aware of the new universe of data that opened for us. Nonetheless, computational power is only a fraction of what information science comprises.

The real revolution comes from Claude Shannon's audacious and masterful thesis (1937) in which he proved that the arrangement of circuit relays could solve logical problems. Shannon's thesis opened the door to the assumption that information can be treated as a physical quantity, the same as mass or energy. The idea that information is at the root of everything has been reinforced by scientists like Norbert Wiener -the author of *Cybernetics: Or Control and Communication in the Animal and the Machine*- Wiener wrote that information, rather than energy, is the central notion in science. In other words, it is not the flow of energy, but the flow of information what defines systems behaviors. Messages are the unifying force behind the natural world we know.

Information is a dynamic concept. According to Roederer (2005), information is embedded in a particular pattern whose purpose is to cause a change of some kind, somewhere. It is transmitted from one entity to another through a physical mechanism - the information processing- to provoke an intended change. The replication of DNA, the lock-and-key behavior of enzymes, and evolution -to name a few instances- can all be understood as messaging processing, or as Roederer stated "life is information at work"(p. 160)

Shannon's information theory delt basically with information itself. In fact, he was more concerned about communication systems. Through his theory, he defined the algorithmic representation of the amount of information embedded in a message and the amount of entropy. He also worked on ways of measuring the degradation of messages during the transmission and storage, which were fundamental concepts for the future development of computers. However, if we take Roederer's definition of information, Shannon's ideas do not attempt to explain the purpose of information. He pointed us, though, to the possibility of replacing physical systems with statistical approaches to make them manageable and understandable for our purposes. His logarithmic information formulas set the basis for computer development and the future advances in the realm of information science. It meant the advent of the digital world.

Information science embraces the concept that every particle of mass, every amount of energy, and every field force we have studied "derives its function, its meaning, its very existence entirely [...] from answers to yes-or-no questions" (Wheeler in Zurek (ed), 1990). In short, messages can be transmitted by strings of bits, in series of 1s and 0s. Most amazingly, a concept like entropy, a fundamental property within a collection of objects, is also applicable in the realm of information which is measured in terms of probabilities as well. According to Seife (2006) “the universe is shaped by the information it contains”(p. 2) Entropy has become a measure of information.

Originally seen as an abstract entity, Information is actually measurable and tangible. It can be represented in many ways like strings of text, flow of electrons, and atoms orientation in magnetic materials to name a few forms. This new way to look at reality aligns not only with the concepts of classical physics but, most importantly with modern theories like Quantum Mechanics. In this realm, Quantics is actually offering the opportunity of developing new generations of devices, which would be able to handle even bigger amounts of data at much faster speeds than our current computers.

How this revolution affects our daily lives, and how knowledge will be handled in repository systems, being them search engines, libraries, or any data management structure, are critical tasks for our future.

The intriguing, and at the same time fascinating aspect of information science is that it comprises not only of concepts related with the study of the natural world around us but also with our societies and networks. Even our personal choices are under scrutiny as we are now questioning if we truly decide by our own or it is simply the influx of information what manages the outcomes. Many discussions have been written on the fact that we, humans, produce information that, when confronted to it, makes us behave in amazingly predictable ways, like a numerical determinism. The core idea of quantum duality of particles and waves, mass and energy fields, are being extrapolated by researchers and compared to the behavior of individuals and the networks around them.

For centuries, reality was understood by the relationship between individuals and their surroundings. Our digital world has shifted our paradigm. Now, reality seems to be rather a faceted, collective construction than an individual interpretation. Like a cocoon, humans are creating an intertwined network where the idea of reality has transformed. The perception of reality seems to be mutating into an interconnecting web around us by which the main element are not the individuals but the whole. “Information is us” (Roederer, p. 1)

To continue this analysis, it might be useful to separate three interrelated concepts: Data, information, and knowledge. Data is basically what we obtain through observation and collection. Information is the organized compilation and summary of the data within its context. Finally, knowledge is the conclusion and awareness we obtain from the analysis of the data collected. It requires an intelligent system to convert data into knowledge. Until now this intelligent system has been our brain. However, the exponential increase of data and the advancement of computers’ technology and Artificial Intelligence (AI) are shifting our mental processes to computational machines to the point that we are increasingly producing and sharing information that we are not even able to read -take QR Codes, for instance-.

Some authors have summarized the history of computers improvements in three basic periods: Tabulation, programming, and AI. In its beginnings, computers were simple machines that served to organize numerical information and perform calculations. With the arrival of microprocessors, it came the programming period. Myriads of applications have been produced to entertain us, to improve our jobs and our productivity. The third stage, the AI, is the one we are experiencing at the present time. AI means that the main focus is on machine learning and it has been envisioned also in three stages. The first one -Narrow AI-, which we are enjoying now, is basically purpose driven and mainly business focused. It includes the current advances in voice computing and smart homes. The next stages, General and Super AI, are expected to occur in the next decades. Future advances in AI target target deep machine intelligence with human brain capabilities as the main focus.

Advances in information science, in conjunction with current extraordinary developments in computational power, have brought to light staggering amounts of human data. From web services that collect information about their users to smart homes and electronic household items; medical devices and surveillance equipment; photographs, videos, and virtual reality; Internet and online social networks… everything, everywhere are producing zillions of bytes of data continuously. This is what we call big data. Ideally, all can be collected, organized and analyzed to provide back extra layers of useful information at many levels: personal, commercial, social, economical, and educational.

Harnessing all this data is the challenge of our era.

With the improvement of our computers came tracking, analytics, conversion, and click-throughs. The increase of computational power has shown patterns and relationships and with them predictive analysis. It was the birth of data mining.

As time passes, the more data that can be captured and processed, the more significance it acquires. Contrary to commodities that tend to lose value with quantity, data is proprietary. The more there is, the higher its value. With statistical analysis, the greater amount of data we have, the closer we are to determine correlations and even causality in any particular situation. With better prediction tools, the better we can reach our goals whatever they may be like, for instance, more sales, better investments, or appropriate vaccines.

In larger data collection environments we find current applications under the denomination of Interactive SMACT technologies which include Social, Mobile, Analytics, Cloud, and IoT. Data generated by mobile systems, IoT, and other networks are collected, aggregated, processed, and analyzed to recognize patterns and extract conclusions. The impact is so profound that data analytics has challenged the belief of our own free will by discovering predictable behavioral patterns. Statistical correlations at large scale have uncovered not only what people may think, but what people will do, which surprisingly is more foresseable than we ever thought possible. In our previous understanding of reality, individuality was the starting premise. However, the current data being gathered shows that each one of us is less surprising and uncommon that we believed. Marketers and sociologists have been able to statistically predict behaviors even in situations in which we thought were completely personal and unique.

Prediction depends on the ability to extract information from the data gathered. This process still requires a human brain. Automation is not good enough yet. Business analytics still requires human insights. Robots are definitely better at calculations than humans, but complex judgements are still human endeavors. Nonetheless, it appears that we are arriving to a pivoting moment. Data is getting unmanageable in size for data scientists and artificial intelligence is getting better at simulating thought processes to turn data into meaning.

Spotting patterns and correlations are becoming a remarkable ability of AI. Nonetheless an additional problem arises. If we let the machine alone to make decisions based on patterns, it might become imperative, as Harkness (2016) suggests to have also an algorithmic accountability system in place, if the analysis in question affects individuals directly. AI should not become a black box. It needs to have a way to correct itself, to self-adjust, and/or to allow input from humans. Programming is not infallible.

Recently, the combination of machine and human thinking has shown interesting results. In some instances, it has been a mutually enriching and learning experience. In 2016, Google's AlphaGo AI machine played Lee Sedol, the world Go champion. In one game, AlphaGo made an inventive move -considered by most a mistake- that ended with a victory for the machine. Sedol on game 4 played also a surprise move letting him win that game. Machine and human managed to surprise and learn from each other. According to Frank (2017) "Sedol [...] became smarter and more sophisticated through his interaction with the machine." (p.142)

Humans, at least for now, manage context better than computers, but the advantage of using algorithms is that we hope to avoid preconceptions and biases to extract conclusions. Netflix uses algorithms to learn what works and what doesn't in its programming all over the world. For instance, contrary to what was intuitively expected, many "Japanese anime films are actually consumed outside Japan." (Frank, p.173) If Netflix had not used its algorithms, it may have lost an important source of revenue. By using only a human mind, within its context, it could have, perhaps, reduced the geography availability of anime programming. Statistical analysis proved better results otherwise.

Big data comes in many forms and it can be useful in multiple ways. Nonetheless, until recently, big data was historically short. In 2002, Google embarked on a project at a grand scale. The purpose was to digitize every book written throughout history, save those of the Alexandria library since all their records were destroyed. At that time, it was estimated that the amount of books ever written in history was in the order of 130 million. With the use of libraries, publishers, and authors Google created an index of books. In 10 years, it had already digitized almost a quarter of them. However, copyrights represented an important factor to consider. Once books were digitized they couldn't be left to be reached freely. Legal considerations needed to be evaluated. Creative ideas call for creative solutions. In 2010, the NGram Viewer was released. Instead of analyzing books as a whole, a mathematical algorithm measures the frequency of words. According to its creators, it is the tool that allows to see "how the past looks like when viewed through a digital lens"(Aiden, 2013, p.23). As the digitized information is text, it is possible to quantify the frequency of a particular word in time, free of any personal contexts. The process relies on computers to see what we can not see. The beauty of this system is that it is long data, hundreds of years in fact. To answer questions, like the one posted by Aiden and Mitchell in their book, when did the United States became one nation? The answer can be charted by comparing the use of the keywords: *the United States **are*** vs. *the United States **is***. The result is easily discovered on the intersection of both graphs.

This ingenious tool, though, seems to work fine if we consider only books, but our modern recorded history is much more than books. There are unpublished texts, magazines, newspapers, audiovisual materials, and objects. Even at a personal level there are plenty of messages, blogs, emails, photos. New solutions -with the required legal and ethical considerations- will need to be developed in order to record a broader spectrum of data without sacrificing anybody's identity or legal rights.

Philosophers like Mayz Vallenilla and many others understood that the current trend of technology will bring humans to a different frontier called the meta-technology. Historically, as Frank affirms "the story of human evolution is in many ways the story of our tools" (p. 135) However, until now, our tools have been "physical" objects, from the first pointed rock to the space shuttle. With software, nanotechnology, and the coming to fruition of quantum devices, our tools are becoming progressively abstract. Our tools are now not objects in strict sense, but mathematical constructions. These modern tools are increasing not only our physical abilities but our mental ones. They are taking learning outside ourselves. For that to occur, machines require instructions to work -programming- and "semantics" to handle information. Meta-technology requires a meta-language, metadata, in order to process information.

Metadata, also called data about data, is a human abstraction, it is a way by which the intrinsic complexity of an object is represented in a simpler form. Pomeranz defines it as an statement about a potentially informative object. As a human tool, it allows people to engage with data by applying chosen categories, "thus allowing us to experience something that otherwise would remain incomprehensible" (Von Müller, 2017, p.). Metadata can be obtained by extracting information from data mining or can be used to classify data in repository systems for future retrieval. In both cases, metadata has three main intrinsic components: what it contains, when and where it is created, and who produces it.

Information is quantifiable. Everything we know is subject to its laws and that includes the fact that data, all data, is multidimensional because it is contextual. The information issue becomes complicated since we have humans in the mix. Data not only has physical or biological qualities. In human systems, it also has mental and social components. This is the conundrum we are living nowadays. Handling data with the help of computational equipment is providing enormous amount of information. Every single activity that we trace, measure, and analyze is providing input on discovering and hopefully solving the widest variety of situations from marketing and retail sales to detecting the onset of diseases or the shift to social unrest. On the other hand, that same multidimensionality creates an enormous complication on how we may retrieve, at a later time, all the knowledge we are currently acquiring.

In information repositories, metadata is not used to capture subjective interpretations of resources, such as relevance, but rather to capture objective features of those resources such as characteristics, origin, use, and organization. Nonetheless, according to Gardner a metadata "always makes a statement about the world, and this statement is subjective in what it includes, what it omits, where it draws its boundaries and in the terms it uses to describe it." (Von Müller, 2017).This point might be particularly relevant if we consider the difficulties in describing visual or musical materials and their subsequent cataloging. Artists, for instance, are usually trying to create original pieces, though at the same time they need to be recognized. Each of their creations have an underlying commonality despite the innovations they may offer. For the public, there is a learning process on recognizing those two aspects. Additionally, the temporal distance between the creator and the spectator adds another layer of cross references and approaches. For modern listeners, used to Glass' music, a piece composed by Bach may convey a completely different meaning than the one for an individual of the XIX century. Humanity learns continuously to handle information and the learning process, originally an individual task, inevitably expands to collective knowledge, but this collective knowledge is also tied to its times. Apple products showed the importance of design and visual appearance on current common tools. In modern times, we have been bombarded with visual and auditory stimuli. Consequently, our visual and aural language is continuously evolving too. What would be the reason for remastering movies decades later they were produced, if it weren't for the fact that our constant information quest is not only a mental, analytical, readable process, but sensorial too.

How we find the information, once it has been organized and cataloged, is a major and increasingly complicated task because it has to be searchable. Searchable means that we need to foresee how it will be looked for in the future. In short, data obtained in the past, has to be cataloged in the present for future reference. Contextualizing implies then, not only the aspects affecting data punctually at the moment is generated, but a time factor foreseeing when it is going to be retrieved and used.

Context affects data because is always interpreted from beginning to end. From the sender to the receiver. From the moment is acquired to the moment is retrieved in the future. Every time someone is researching information it does it on the premise of what this person already knows, it is based on his/her own assumptions. Every time raw data is obtained has to be processed to extract understanding and conclusions, but hose conclusions are done by data scientists that also carry assumptions and mental paradigms. Every time a system is set in place to catch raw data, it has been done under premises relative to that present moment.

Inevitably, all what we know occurs within a cultural context. Facts, or data about facts, need to be comprehended and analyzed within context. As Buckland (2017) mentions, "our knowledge, modes of communication and ways of reasoning are culturally situated in our personal small world" (p. 57).

No matter if we are talking about a discrete amount to content, like the ones we may find at a small library branch, or a huge data generating network like Facebook or Blogger, the imminent difficulty we are facing is an increasing need for faster, efficient ways to classify and analyze data in order to be able to extract exactly what we are looking for at a later time.

The issue is not only finding a previously identified document. The real challenge is to retrieve the most suitable one which it is not known yet. For this, we commonly follow a 2 steps process. The first one is the creation of the collection of documents and the second one is having the tools in place that will allow the search within the collection. For the latter, we require a productive and efficient way to generate metadata so it can serve its two intrinsic critical purposes: description and search.

Metadata contains, first of all, a name. Naming uses a language, and consequently it is a cultural activity. This is the first difficulty. The process of naming needs a protocol, a particular notation, but notation is defined before naming. Inherently, naming is attached to the past. It is obsolete from the very beginning. For future searchers, the name might not have the same meaning neither provide the expected results. A search for a simple word as "windows" would have produced completely different results few decades ago compared to what we find today.

If naming is complicated, description is even more. Since for a search to be successful relies on a suitable description, the intricacies of language evolution, time, and contextual situations will affect the process. The duality of a defined set of rules and the ever changing words, places metadata writing on a difficult conundrum. Done by humans, social agendas, biases, and preconceptions affect descriptions. On the other hand, if metadata is restricted by notation protocols, it disregards language evolution and nuances. Ngram has shown how much words, and their use, change in time. As explained above, a simple difference of singular or plural can make one nation.

It is inevitable. The world changes so the names and descriptions are compelled to change. However, that is not all. The purpose of information repositories (e.g. a library) is to describe for the future, not only in linguistic ways, but also in content ways. In other words, semantics is not the only variable, our understanding and knowledge as individuals, as well as our collective progress, should also be considered. For instance, once an individual performs one search, his second one is inevitably affected by the first one. There has been a learning process in the simple fact of search that ideally could be reflected on the next one.

For that matter, artificial intelligence has become the next indispensable step in our race to keep up with ourselves. We are in a loop in which AI needs to enhance and revamp constantly. We are continuing increasing our data consumption and our repositories need to perform accordingly. Search engines like Google, for instance, are learning to know us so they can even finish our own sentences when we are searching for information. And yet, computer learning brings another difficulty. If the system only learns from the user, it will continuously reduce the universe of results to resemble the individual perceived need. Relevance becomes then dynamic, situational, and variable, but also could become limited and restrictive.

Using the individual experience as a learning tool might degenerate into what is being known, in the digital world, as balkanization of knowledge. Information sources, like libraries or search engines, are based on servicing humans to provide the information required and to encourage their acquisition of information, ideally on a beneficial way, and to transform it into personal knowledge. Nonetheless, diminishing their universe of results to similar patterns in the user's minds is it really beneficial? How can we question our own ideas if we are not challenged by others? As Edwards (2015) mentions in regards to the digital effects "we lose the benefit not just of being able to influence others by our personal impression but also of possible influence by others" (p. 55) This is the delicate, complicated balance in which information systems recognize themselves.

We still have a long way to go to see the egalitarian, multiverse, independent voices and views that utopians believed we will reach. Not because those don't exist, or because they don't express themselves, but because they are not talking to each other. Individuals get easily trapped inside information bubbles were they keep corroborating their own, or their perceived equals, statements.

Exposure foments learning. Without education experiences, without information within context, societies are poorly educated which translates in diminishing opportunities. "Poorly educated societies tend to be poor societies" (Edwards, 2015, p.61)

As individuals, creativity and reflexion should be part of our acquisition of knowledge if we really want to move forward toward innovation on every aspect of our lives. Here perhaps lies the core of information science, to balance the machine possibilities and capabilities with the socio-cultural element embedded on every human activity. Powerful search engines like Google, Wolfram Computational Knowledge, Cellular Automata, and other technologies are developing alternatives for our constant quest to conquer complex realities. However, the road is challenging because many factors are involved. From ethical, and legal issues to privacy, and security. From personal interests, to businesses and governments. Our society, as we know it, is at an undergoing gigantic transformation.

With new technologies, systems increasingly rely on machines to read documents, save data, and allow searches. Our own world is too complex to be handled and comprehended by individuals and in order to expand our reach, our tools need to extend our capabilities.

There are tools being created to facilitate the task of cataloging. Reliable semi automatic metadata generation, for instance, is already in place even though it is still a work in progress. Its main problem, though, is that it does not respond the same way to different materials to be cataloged. (Park, 2015)

Many other issues are emerging with the advent of AI. Fake videos, digital manipulations, face digitizations, and morphing are just a few of them. They are casting a shadow on how to choose the materials we preserve. What information should be kept? Is it possible to recognize all the real ones from the manipulated ones? How to discriminate an edited copy from an original? Should we keep everything?

The problem is that the same technology that facilitates the work of the information systems is the one that can also interfere and distort our repositories. It will require an army of well-trained, well-versed, technology-savvy catalogers and search developers to make sure we preserve the information as reliable and accessible as possible for future use.

References

Aiden, E., & Michel, J. (2013). *Uncharted : big data as a lens on human culture*. New York, N.Y.,USA : Riverhead Books

Akl, S. G. (2017). Information and Computation: The Essence of It All. *International journal of unconventional computing*, (3). 187.

Buckland, M. (2017). *Information and society*. Cambridge, Massachusetts ; London, England : The MIT Press

Edwards, A. (2015) *Digital Is destroying everything: What the tech giants won't tell you about how robots, big data and algorithms are radically remaking your future*. London, UK: Rowman & Littlefield

Frank, M., Roehrig, P., & Pring, B. (2017). *What to do when machines do everything : how to get ahead in a world of AI, algorithms, bots, and big data*. Hoboken, N.J.,USA: Wiley.

Gartner, R. (2016), *From Information to Knowledge: A Beginner's Guide to Metadata*, Heidelberg: Springer.

Gimbel, S. (2015). *Redefining Reality: The intellectual implications of modern science*. The Great Courses.

Gleick, J. (2011). *The Information*. New York, N.Y., U.S.A: Pantheon Books.

Harkness, T. (2016). *Big data. Does size matter?* Croydon, UK: Bloomsbury Sigma.

Hwang, K., & Chen, M. (2017) *Big-Data Analytics*. Hoboken, N.J.USA: John Wiley & Sons Ltd.

Intner, S. S., Lazinger, S. S., & Weihs, J. (2006). *Metadata and its impact on libraries*. Westport, Conn. : Libraries Unlimited.

Mayz, E., (1993) *Fundamentos de la Metatecnica*. Spain:Gedisa

Park, J., & Brenza, A. (2015). Evaluation of semi-automatic metadata generation tools: a survey of the current state of the art. *Information Technology And Libraries*, (3), 22. doi:10.6017/ital.v34i3.5889

Pierce, J. R. (1980). *An introduction to information theory : symbols, signals & noise*. New York, N.Y.,USA: Dover Publications

Pomerantz, J. (2015) *Metadata.* Cambridge, MA., USA: MIT Press

Roederer, J. G. (2005). *Information and Its Role in Nature*. Berlin: Springer.

Seife, C. (2006). *Decoding the universe : how the new science of information is explaining everything in the cosmos, from our brains to black holes*. New York, N.Y.,USA: Viking.

Shannon, C. (1938) A Symbolic Analysis of Relay and Switching Circuits. *Transactions American Institute of Electrical Engineer*s, (57) 38-80

Von Müller, J. (2017). Metadata: New perspectives on Aby Warburg's 'critical iconology'. *Philosophy Of Photography*, 8(1/2), 171-182.

Wiener, N. (1967). *Cybernetics; or, Control and communication in the animal and the machine*. New York,N.Y., USA: M.I.T. Press.

Zurek, WH (ed.) (1990) *Complexity, entropy, and the physics of information : the proceedings of the 1988 Workshop on Complexity, Entropy, and the Physics of Information, held May-June, 1989 in Santa Fe, New Mexico*, Redwood City, CA : Addison-Wesley Pub. Co.