

Using the information and communications technology data deluge from a semantic perspective of a dynamic challenge: What to learn and what to ignore?

-Part 2-



Prof.Eng.Ph.D. Victor GREU

Abstract

The paper analyses the influences of the Data Deluge (DD), as huge flows of data leveraged or created, in all activity field, by the complex proliferation and exponential development of Information and Communications Technology (ICT), as main driving factor of the progress of the Information society (IS) toward Knowledge Based Society (KBS). As a consequence, the paper further approached the dynamic processes **where the ICT generates the data**, by the complex symbiosis of ICT with humans.

In a systemic approach, the paper shortly analyses the revolutionary impact of DD on science, but also the way the most of DD is generated (including the challenge of the technologic progress necessary for accomplish this task), resulting that the DD has leveraged a new reality, called eScience, considered the fourth paradigm, after experiment, theory and simulations, where scientists were no longer interacting directly with the phenomena, which not only open the exploration of (inaccessible) new fields, but provide the base of data-intensive science - one of the mechanisms where the World spiral development is produced using **the multiplication force of ICT/DD**. The concrete features of this new era (visible at CERN, climate change forecasting or critical national infrastructures), represent **fundamental trends in solving the most critical challenges of DD** and include Big data analytics, real time processing/storage on site and the integration of ICT advances like Internet of Things (IoT), artificial intelligence (AI), Cloud, Edge, Fog etc. Among these trends, the new phase of AI is now the ICT most important hype, remarkable by machine learning (ML), deep learning (DL) and further the emerging cognitive computing, but **their overall results are also highly dependent on** human intelligence (HI) for delivering the optimal knowledge refining. The long road from data to knowledge or decisions has many steps, depending on (the amount of) data, but also on specific field/algorithm and humans, as the highly performant applications need usually (for training) sample data with specific content (not only simple/unstructured ... more data), which often ask human supervision, although AI/ML not only tend to get closer to HI, but for specific task to replace humans.

Another side of DD challenges is about the most complex and complicate issues of extracting and especially **evaluating knowledge gain** in the diverse processes involving the design and use of DD/AI/ML at Earth scale. This is about how to **manage DD evolution, in each main field of applications**, in order to keep them efficacious and efficient.

A preliminary conclusion is that such approach could lead to partial reasonable solutions, but it eludes just the most important and difficult issue, from our point of view, i.e. the global effects (influences) of ICT/DD exponential evolution at Earth ecosystem scale. Even mathematically, the subsystems could be optimized by some criteria, but at system level not only it could appear a suboptimal solution, but we could expect to considerable vulnerabilities and risks. Besides the huge benefits, the exponential ICT development (without excluding other connected technologies) could generate at Earth scale less desired consequences like carbon footprint, human dependences, wastes pollution etc., in the general World context of climate changes, resources fading, clean environment, social unbalances and so on. The most relevant case is that DD leverages many applications for climate changes forecasts and this is a benefic influence, but here it could be a vicious circle, because these applications represent only a little part of DD sphere and many others could be less benefic or knowledge providing, like streaming to much video content for entertainment or games, while **all** applications from sphere contribute to carbon footprint. The analysis should be extended this way to **all** applications, i.e. it is necessary to evaluate them all, counting, case by case, at two levels: first at local level, the specific **benefits and challenges** and second, at global level, the complex connections by which they could be added or influence others in **superposition (less benefic) processes or consequences**.

*As a consequence, the most complicate and difficult problem of DD/ICT exponential evolutions management is how to obtain, using both DD/ML/AI/ICT and HI resources, that refined knowledge which could provide optimal solutions to every relevant phase of those evolutions, i.e. how to save the Earth and humankind from irreversible consequences using the most powerful tools of science and technology (HI/AI). More than these, the difficulty of this global problem is increasing with the level of complexity and with the speed of changes DD/ICT generate to IS/KBS. In this global problem, the engineers and other specialists involved in developing ICT systems, products and services have a prominent role (unfortunately not always decisive one). In fact, the fundamental problem of optimally refining the knowledge often starts at the technical designers, although **the opportunity of developing big projects should involve many levels and different areas specialists, in order to get a multicriterial optimization at global scale.** As opportunity means also what knowledge must be refined, this problem is similar with the (university) conscious professor's dilemma, we all experience (desirably) every year at the beginning of school when we decide what to erase and what to introduce in our course, and generally **when** a course, book or knowledge become obsolete.*

The final conclusion is that in our ever-changing DD days, it is almost impossible to precisely know when and how knowledge must be refined (a relative and approximative decision), but this does not mean that humankind have to ignore these challenges, on the contrary, they have to continuously look for getting as much as possible close to optimum, by the available updated data/information and resources, using AI/HI with desirable wisdom solutions.

Keywords: Data Deluge, Big Data, machine learning, artificial intelligence, learning algorithms, human intelligence, Internet of Things, eScience, data-intensive science, computer algorithms, computing infrastructure, climate changes

JEL Classification: L63; L86; M15; O31; O33

***“Information is not knowledge.”
- Albert Einstein***

1. How Data Deluge could influence science, technology and people?

The history is full of evidences of the continuous increase of the humankind preoccupation to create material forms for spreading information and refining knowledge, even in the ancient times when terms like data and information were not used like today.

Before a further development of this idea, it is interesting to observe that in the same old times, instead of data or information, knowledge and wisdom were used, but with meanings very similar with their actual significances, which shows their perennial importance.

Coming back to data and history, the evidences show that the mind creations were transmitted by written supports (like **stone and paper**) or other artefacts like paintings or even sculptures.

It is easy to observe that all these formats had a material support, even in the impressing case of music, where sound was a case when the information had no material support till the invention of musical notes, although the speech and poetry had similar evolutions before writing being invented.

The deep meaning of these examples could be, by our opinion, two fold, as when it was no usual material support all these mind creations were temporary received and stored in other ... people's minds.

It is impressing to also observe that, although people did not know, they had to store data in their brains memory cells, just like we do today with all storage technologies (from tapes to hard/optical discs or solid state memories).

The other meaning could be the way how these two storage supports (brain cells and modern magnetic, optical or other memory cells) reflect the spiral evolution.

More than these, the actual importance of such observations is based on the fact that, in spite of the amazing progress of Information and Communications Technology (ICT), as

main driving factor of the progress of the Information society (IS) toward Knowledge Based Society (KBS), the natural model of human brain is still a source of inspiration and innovation.

In fact, as we have already mentioned [21], the nature research of millions of years generally is still an inspiration for the apogee semiconductor technologies of our days, which is just the most prominent example, among other scientific or technical fields of IS/KBS.

Now we have just arrived to the main goal of our paper, intended to analyse how these ICT evolutions, including the actual phase of **Data Deluge** (DD), could influence the humankind and Earth evolution, by leveraging the refined knowledge creation, necessary for the actual and future complex context of IS/KBS.

Essentially, DD consists in the huge flows of data leveraged or created, in all activity field, by the complex proliferation and exponential development of ICT in the context of IS/KBS.

Concerning Data Deluge definition, there are many points of view, but its main association, in the technological field, is with **Big Data** (BD), which includes specialized equipment and services (hardware/software) intended to optimally process the huge amount of data provided by Data Deluge.

Analysing DD influence at the Earth ecosystem scale is difficult because of the complicate and complex processes we have to consider when approaching the Data Deluge and generally the exponential ICT development and consequences, which could not include only benefic ones, but, at least, these challenges should be timely approached and continued [16][3].

Coming back to the **stone and paper information**, one could ask: People always wanted to share their thoughts and experience in order to improve their lives, what is now the main difference?

We consider that this is a good question because the perpetual knowledge dilemma (**What to learn and what to ignore?**) has not changed much, but in our days the difference is on the road from **stone** to ... **ICT electronic support**. Here, our Einstein quote, “*Information is not knowledge*”, seems to be more than relevant and full of truth.

The essential observation is that only after understanding the deep roots of this difference, we could forecast the ways Data Deluge influence should be approached and optimized in order to leverage the actual/appropriate refined knowledge creation, in a complicate global process of check, act and balance, just because **ICT/DD and actual knowledge are dependent variables in global random processes**.

Apparently the support is only a part of the complex proliferation of data, but this view could change if we understand that **ICT electronic support** is more than “support”, as it inherently includes the **huge sphere** where ICT (and then Data Deluge) is spreading its influence (by products and services), with **exponential pace of change**, on whole Earth ecosystem, becoming the main driving factor of the progress of the IS toward KBS.

More than this planetary sphere, we have to carefully analyse the dynamic processes **where the ICT generates the data**, by the complex symbiosis of ICT with humans.

This way we have just arrived to the **core** of the mentioned complex processes at the World scale, because of ICT penetration to all activity fields and beyond, influencing Earth environment and eventually humankind evolution ... everyday with an amazing speed.

The complications come mainly from these global and exponential changes, which are more and more difficult to control and to be estimated regarding all their (positive or negative) consequences.

If all the above assertions could appear as subjective and not sustainable, it is useful to analyse some relevant examples, from the above **sphere** and **core**, which reflect the complicate and dynamic data/information/knowledge generation processes in IS/KBS, although these examples could be just the (ICT/IS/KBS) **iceberg tip**, as we already have mentioned [19]. In technological terms, we could also say, for example, that Big Data is, metaphorically, only the **iceberg tip** of DD, as it will further result.

Although even the science is in a flourishing and challenging époque, due, among others, just to ICT/DD, as any deep approach should start with a scientific point of view, a first example is just a global image of DD impact on this basic field [1]:

“Since at least Newton’s laws of motion in the 17th century, scientists have recognized experimental and theoretical science as the basic research paradigms for understanding nature. In recent decades, computer simulations have become an essential third paradigm: a standard tool for scientists to explore domains that are inaccessible to theory and experiment, such as the evolution of the universe, car passenger crash testing, and predicting climate change. As simulations and experiments yield ever more data, a fourth paradigm is emerging, consisting of the techniques and technologies needed to perform data-intensive science”

The essence of this example is revealing the paradigm of *computer simulations*, which not only open the exploration of (*inaccessible*) new fields, but provide the base for generating the new *fourth paradigm*, i.e. *data-intensive science* - one of the mechanisms where the spiral development is produced using **the multiplication force of ICT/DD**:

“For example, new types of computer clusters are emerging that are optimized for data movement and analysis rather than computing, while in astronomy and other sciences, integrated data systems allow data analysis and storage on site instead of requiring download of large amounts of data”

In addition to the new fundamental approach, *data analysis and storage on site*, also applied at CERN and mentioned in [10][3], we must observe the connection of DD/ICT with one of the prominent challenges of our days: *climate change* - a crucial problem that Earth ecosystem is facing with and we all have to continuously debate.

It is clear now not only the revolutionary impact of DD on science, but also the way the most of DD is generated (including the challenge of the technologic progress necessary for accomplish this task). An argument for this estimate is also well documented in [3], where CERN scientific centre data generation capabilities are detailed presented.

For confirming the above assertion and our observations, another point of view is all relevant [2]:

”In 2007 Jim Gray preached about the effects of the Data Deluge in the sciences (Hey, Tansley, and Tolle 2009). Whereas experimental and theoretical paradigms originally led science, some natural phenomena were not easily addressed by analytical models. In this scenario, computational simulation arose as a new paradigm enabling scientists to deal with these complex phenomena. Simulation produced increasing amounts of data, particularly from the use of advanced exploration instruments (large-scale telescopes, particle colliders,

etc.) In this scenario, scientists were no longer interacting directly with the phenomena, but used powerful computational configurations to analyse the data gathered from simulations or captured by instruments. Sky maps built from the Sloan Digital Sky Survey observations, or the evidences found about the Higgs Boson are just two successful stories of just another paradigm, what Gray called the fourth paradigm: the eScience.”

Here it is worth to notice that the DD leveraged a new reality, where scientists were no longer interacting directly with the phenomenon, (the same fourth paradigm) which is also called eScience in [11].

As the dynamic challenges we have mentioned changes everything with an amazing pace, even this estimation tends to become obsolete because another huge data flow, generated by Internet of Things (IoT), could surpass in the future the scientific one [4]:

“Critical national infrastructures, such as electricity grids, water and waste water systems, etc., also use IoT for system health monitoring to achieve the goal of early detection of faults in the target system in order to exercise timely mitigation plans... A large volume of data is collected from a large number of sensors and these data streams flow continuously. Thus, data storage and data analysis are major issues to be considered. Big data analytics have been widely used in recent years to analyse the data collected from IoT systems to achieve better fault detection and diagnosis ... for the target system. Due to the computation intensive nature of big data processing, the analyses generally need to be performed in the cloud. However, channelling the data to the cloud may have significant communication latency. Also, a lot of the fault detection and diagnosis tasks need to be performed in real time. In response to such conflicts between timeliness and resource constraints, the new computing infrastructure that integrates IoT, Edge, Fog, and the Cloud have been developed. Preliminary computations that are required for timely decisions can be done in the nearby Edge gateways and Fog LANs (local area networks). Integrated analysis can be left to the Cloud and the more sophisticated and/or more accurate analysis results can then be channelled back to the Edge and IoT devices to provide improved decision making or to support changing trends.”

Far from adding only the news of *Big data analytics* (as one of the branches of the Big Data), this example is the very relevant for showing (still only) a (deep/core) part of the DD/ICT **iceberg**, by pointing the ICT/IoT/DD crucial role in *critical national infrastructures*.

On the other hand (i.e. another deep/core part of the DD/ICT **iceberg**), we can see now the *sophisticated* challenges (as *conflicts between timeliness and resource constraints*) the *real time* processing of such huge data flows rise for the implementation of *the new computing infrastructure that integrates IoT, Edge, Fog, and the Cloud*.

It is important to notice that here, just like in the previous example, **on site** approach, involving new performant technologies (*IoT, Edge, Fog, and the Cloud*), is **a fundamental trend in solving the most critical challenge of DD**.

Obviously, although these examples are very relevant, the list could be continued, but now it became clear that the mentioned DD/ICT **iceberg** should be further analysed not only in the sphere dimensions, but especially in the complexity (depth) of the associated processes involving the hardest part, i.e. **where and how the refined knowledge could be generated**.

2. Learning from data – a dynamic challenge: from machine learning to knowledge

Everyday and everywhere we can see the ICT products and services spreading with an amazing pace, bringing benefits and surprising us with their (most of them pleasant) consequences.

This simple observation could mean, in the same time, that our further analysis could be much extended in the direction of DD direct influences, but for a sustainable progress of ICT in IS/KBS it is mandatory to see the whole picture, i.e. including the role of humans, especially of the human intelligence (HI), in the complicate and complex processes associated with ICT/DD exponential development, which actually has another emergent hype: the new phase of artificial intelligence (AI) performances.

In short view, the actual AI phase is remarkable by machine learning (ML), deep learning (DL) and further the emerging cognitive computing (CC), as we have already approached them [6][9][13].

The link of DD/BD with AI/ML is perhaps one of the actual most prolific, dynamic and complex process of ICT/IS/KBS, as it tends to reach incredible level of performance, pointing the HI, but it also produces a huge range of applications – all including the analyse and processing of large amounts of data.

The long road from data to knowledge or decisions has many steps, depending on ... (the amount of) data, but also on specific field/algorithm and ... humans, all of large diversity, as it is well presented in [5]:

<<The collection of “Big Data” and the expansion of the Internet of Things (IoT) has made a perfect environment for new AI applications and services to grow. Applications based on AI are already visible in healthcare diagnostics, targeted treatment, transportation, public safety, service robots, education and entertainment, but will be applied in more fields in the coming years. Together with the Internet, AI changes the way we experience the world and has the potential to be a new engine for economic growth...

Although artificial intelligence evokes thoughts of science fiction, artificial intelligence already has many uses today, for example:

- Email filtering: Email services use artificial intelligence to filter incoming emails. Users can train their spam filters by marking emails as “spam”.*
- Personalization: Online services use artificial intelligence to personalize your experience. Services, like Amazon or Netflix, “learn” from your previous purchases and the purchases of other users in order to recommend relevant content for you.*
- Fraud detection: Banks use artificial intelligence to determine if there is strange activity on your account. Unexpected activity, such as foreign transactions, could be flagged by the algorithm.*
- Speech recognition: Applications use artificial intelligence to optimize speech recognition functions. Examples include intelligent personal assistants, e.g. Amazon’s “Alexa” or Apple’s “Siri” ...>>*

Beyond the impressing diversity of the potential impact, it is worth to observe the basic dependence on *Internet* and also the crucial fact that *AI changes the way we experience the world and has the potential to be a new engine for economic growth*, although when

comes the *personal assistants* issue (and similar services) we have to point the interfaces where, more than otherwise, the efficacy and efficiency of products and services, referring to the potential knowledge gain, will strongly also depend on ... humans sitting on both sides (designer/user):

“...The Internet Society recognizes that understanding the opportunities and challenges associated with AI is critical to developing an Internet that people trust. This is particularly important as the Internet is key for the technology behind AI and is the main platform for its deployment, including significant new means of interacting with the network...”

Further, the intimate mechanism of processing data and delivering knowledge will be determined by ML/DL algorithms:

“...Algorithms, developed by programmers to instruct computers in new tasks, are the building blocks of the advanced digital world we see today. Computer algorithms organize enormous amounts of data into information and services, based on certain instructions and rules. It’s an important concept to understand, because in machine learning, learning algorithms – not computer programmers – create the rules. Instead of programming the computer every step of the way, this approach gives the computer instructions that allow it to learn from data without new step-by-step instructions by the programmer. This means computers can be used for new, complicated tasks that could not be manually programmed...”

Perhaps the most important observation here is the fact that AI/ML are highly useful for new, complicated tasks that could not be manually programmed, i.e. AI/ML not only tend to get closer to HI, but for specific task to replace humans.

Still, we have to point that this trend has a very long road to walk because of the critical role of data:

“...The basic process of machine learning is to give training data to a learning algorithm. The learning algorithm then generates a new set of rules, based on inferences from the data. This is in essence generating a new algorithm, formally referred to as the machine learning model. By using different training data, the same learning algorithm could be used to generate different models. For example, the same type of learning algorithm could be used to teach the computer how to translate languages or predict the stock market.

Inferring new instructions from data is the core strength of machine learning. It also highlights the critical role of data: the more data available to train the algorithm, the more it learns. In fact, many recent advances in AI have not been due to radical innovations in learning algorithms, but rather by the enormous amount of data enabled by the Internet.”

In fact, as the highly performant applications need usually sample data with specific content (not only simple/unstructured... *more data*), which often ask human supervision (in general ML could be unsupervised or supervised) or even creation, **the overall results are also highly dependent on humans**, in many ways which should be further analysed.

At a global scale, in the last years, AI/ML applications which benefit from DD/BD are in aggressive expansion, even beyond above mentioned scientific fields, as, naturally, the economy is the largest domain of human activity, where ICT/IS/KBS are based [7]:

“IBM’s recent announcements of three new services based in Watson technology make it clear that there is pressure in the enterprise software space to incorporate

new technologies, both in hardware and software, in order to keep pace with modern business. It seems we are approaching another turning point in technology where many concepts that were previously limited to academic research or very narrow industry niches are now being considered for mainstream enterprise software applications.

Machine learning, along with many other disciplines within the field of artificial intelligence and cognitive systems, is gaining popularity, and it may in the not so distant future have a colossal impact on the software industry...”

A deeper vision of ML and of subsequent processes involved in data analysis is further present as:

“...Basically, a machine learning system learns by experience; that is, based on specific training, the system will be able to make generalizations based on its exposition to a number of cases and then be able to perform actions after new or unforeseen events. The discipline of machine learning also incorporates other data analysis disciplines, ranging from predictive analytics and data mining to pattern recognition. And a variety of specific algorithms are used for this purpose, frequently organized in taxonomies, these algorithms can be used depending on the type of input required ...”

As we have mentioned above, in order to achieve higher performances or very special applications, besides other data analysis disciplines, AI/ML processes are human supervised and this often supposes contributions of experts or highly trained personnel (for building expert systems):

<<...One of the more important applications of machine learning is to automate the acquisition of knowledge bases used by so-called expert systems, systems that aims to emulate the decision making process of human expertise in a field. But the scope of its application has been growing. In Applications of Machine Learning and Rule Induction, Langley and Simon review some major paradigms for machine learning scenarios, all based on a very important premise: To improve performance on some task, and the general approach involves finding and exploiting regularities in training data. The major approaches include using neural networks, case-based learning, genetic algorithms, rule induction, and analytic learning. While in the past they were applied independently, in recent times these paradigms or models are being used in a hybrid fashion, closing the boundaries between them and enabling the development of more effective models. The combination of analytic methods can ensure effective and repeatable and reliable results, a required component for practical usage in mainstream business and industry solutions. >>

It seems now that the DD/AI/ML picture is more and more enlarging and, naturally, this could bring inherent complications and challenges, in design, development and use [15][22].

Let just observe, from the above, only the price of *effective and repeatable and reliable results*, as the apparent benefits need now that independent *paradigms or models* to be used in a *hybrid fashion, closing the boundaries between them and enabling the development of more effective models*.

These details still show only the scientific and technological challenges of building the most advanced AI/ML systems which use DD to extract knowledge or provide assistance for decisions in management. To the list we could add storage, capacity and latency problems the use of DD have to face [14] [20] [17] [18].

Another side of DD challenges is about the most complex and complicate issues of extracting and especially **evaluating knowledge gain** in the diverse processes involving the design and use of DD/AI/ML at Earth scale.

Although this **face** of DD seems (and almost is) very difficult to approach, the paper space could still allow a side (essential) of the problem, which is about how to manage DD evolution, in each main field of applications, in order to keep them efficacious and efficient.

We have to admit, though, that such approach could lead to partial reasonable solutions, but it eludes just the most important and difficult issue, from our point of view, i.e. the global effects (influences) of ICT/DD exponential evolution at Earth ecosystem scale.

As a matter of fact, even mathematically, the subsystems could be optimized by some criteria, but at system level not only it could appear a suboptimal solution, but we could expect to considerable vulnerabilities and risks.

To be more clear, we just recall what repeatedly we have presented [16] [23], that besides the huge benefits, the exponential ICT development (without excluding other connected technologies) could generate at Earth scale less desired consequences like carbon footprint, human dependences, wastes pollution etc., in the general World context of climate changes, resources fading, clean environment, social unbalances and so on.

Of course, one could simple ask: Which is the connection of these consequences with the knowledge extraction from DD?

The most relevant case is that DD leverages many applications for climate changes forecasts and this is a benefic influence, but here we obviously have a vicious circle, because these applications represent only a little part of DD sphere and many others could be less benefic or knowledge providing, like streaming to much video content for entertainment or games, while **all** applications from sphere contribute to carbon footprint.

The analysis could be extended this way to **all** applications, i.e. it is necessary to evaluate them all, counting, case by case, at two levels: first at local level, the specific **benefits and challenges** and second, at global level, the complex connections by which they could be added or influence others in **superposition (less benefic) processes or consequences**.

This way we have just arrived to the core of the most complicate and difficult problem of DD/ICT exponential evolutions management: **how to obtain, using both DD/ML/AI/ICT and HI resources, that refined knowledge which could provide optimal solutions to every relevant phase of those evolutions, i.e. how to save the Earth and humankind from irreversible consequences using the most powerful tools of science and technology (HI/AI)?**

Unfortunately, **this complicate question does not have a single (true) answer**, as we already approached before [3] [9], but people surely have to continuously try to find partially optimal solutions.

More than these, the difficulty of this global problem is increasing with the level of complexity and with the speed of changes DD/ICT generate to IS/KBS.

On the other side, it is sure that the engineers and other specialists involved in developing ICT systems, products and services have a prominent role (unfortunately not always decisive one).

In fact, the fundamental problem of optimally refining the knowledge often starts at the technical designers, although **the opportunity of developing big projects should involve many levels and different areas specialists, in order to get a multicriterial optimization at global scale.**

Still, the next example is very relevant even beyond the engineering level, as it is presented in [8]:

“I was listening recently to some engineering graduates talking about their current research efforts. As I was forcibly immersed in the minutiae of opaque mathematics, the thought came to me that this was really difficult work for potentially small gains. But on the heels of that came another thought: These engineers were really skilled. I was greatly impressed by their depth of knowledge and facility with analysis. How did they learn so much in their few years of university training? After all, there is so much more to know now than there used to be, and every year it gets ever more overwhelming ...”

As usually, most of situations in ITC seem to have only impressing good parts, like above, *their depth of knowledge*, but besides the fact that *every year it gets ever more overwhelming*, knowledge meets the actual high dynamic challenge, we repeatedly mentioned before [16][9], as how to update it, globally, by optimally refining it, which is also further approached:

*“...One explanation is that, as new knowledge accumulates, some old knowledge becomes irrelevant and falls off the knowledge stack. Almost all the college course work I took long ago is now useless in itself, although what remains is an engineering mind-set and a mathematical grounding. Perhaps every course should have a sell-by date. Indeed, in retrospect I now realize that a number of the courses I took were already well past their sell-by dates when I took them. I remember too when the technical library in the lab where I worked was shut down and all the books were offered free to the staff. Almost all of them went unclaimed; no one wanted them. The problem is that we never quite know **when** a particular course or book will become obsolete...”*

Considering that **opportunity** means also **what** knowledge must be refined, we apparently arrived to the (university) conscious professor's dilemma, we all experience (desirably) every year at the beginning of school: What to erase and what to introduce in our course?

As a matter of fact, the example shows much more than this, first, as the constructive and optimistic part, we have to notice that *what remains is an engineering mind-set and a mathematical grounding* - which, by the way, is not little!

A second and the most challenging part is *that we never quite know **when** a particular course or book will become obsolete.*

By our opinion, this way we could only agree that, in our ever-changing DD days, it is almost impossible to precisely know when and how knowledge must be refined (a relative and approximative decision), but this does not mean that humankind have to ignore these challenges, on the contrary, we have to continuously look for getting as much as possible close to optimum, by the available updated data/information and resources, using AI/HI with desirable wisdom solutions.

3. Conclusions

Data Deluge (DD) essentially consists in the huge flows of data leveraged or created, in all activity field, by the complex proliferation and exponential development of Information and Communications Technology (ICT), as main driving factor of the progress of the Information society (IS) toward Knowledge Based Society (KBS).

For this complex explosion of data, **ICT electronic support** is more than “support”, as it inherently includes the **huge sphere** where ICT (and then Data Deluge) is spreading its influence (by products and services), with **exponential pace of change**, on whole Earth ecosystem. As a consequence, the paper further approached the dynamic processes **where the ICT generates the data**, by the complex symbiosis of ICT with humans.

In a systemic approach, we have shortly analysed the revolutionary impact of DD on science, but also the way the most of DD is generated (including the challenge of the technologic progress necessary for accomplish this task), resulting that the DD has leveraged a new reality, called *eScience*, considered the *fourth paradigm*, after experiment, theory and simulations, where *scientists were no longer interacting directly with the phenomena*, which not only open the exploration of (*inaccessible*) new fields, but provide the base of *data-intensive science* - one of the mechanisms where the World spiral development is produced using **the multiplication force of ICT/DD**.

The concrete features of this new era (visible at CERN, climate change forecasting or *critical national infrastructures*), represent **fundamental trends in solving the most critical challenges of DD** and include Big data analytics, real time processing/storage on site and the integration of ICT advances like Internet of Things (IoT), artificial intelligence (AI), Cloud, Edge, Fog etc.

Among these trends, the new phase of AI is now the ICT most important hype, remarkable by machine learning (ML), deep learning (DL) and further the emerging cognitive computing, but **their overall results are also highly dependent on** human intelligence (HI) for delivering the optimal knowledge refining.

The long road from data to knowledge or decisions has many steps, depending on (the amount of) data, but also on specific field/algorithm and humans, as the highly performant applications need usually (for training) sample data with specific content (not only simple/unstructured ... *more data*), which often ask human supervision, although AI/ML not only tend to get closer to HI, but for specific task to replace humans.

Another side of DD challenges is about the most complex and complicate issues of extracting and especially **evaluating knowledge gain** in the diverse processes involving the design and use of DD/AI/ML at Earth scale. This is about how to **manage DD evolution, in each main field of applications**, in order to keep them efficacious and efficient.

A preliminary conclusion is that such approach could lead to partial reasonable solutions, but it eludes just the most important and difficult issue, from our point of view, i.e. the global effects (influences) of ICT/DD exponential evolution at Earth ecosystem scale. Even mathematically, the subsystems could be optimized by some criteria, but at system level not only it could appear a suboptimal solution, but we could expect to considerable vulnerabilities and risks. Besides the huge benefits, the exponential ICT development (without excluding other connected technologies) could generate at Earth scale less desired consequences like carbon footprint, human dependences, wastes pollution etc., in the general

World context of climate changes, resources fading, clean environment, social unbalances and so on.

The most relevant case is that DD leverages many applications for climate changes forecasts and this is a benefic influence, but here we obviously have a vicious circle, because these applications represent only a little part of DD sphere and many others could be less benefic or knowledge providing, like streaming to much video content for entertainment or games, while **all** applications from sphere contribute to carbon footprint.

The analysis should be extended this way to **all** applications, i.e. it is necessary to evaluate them all, counting, case by case, at two levels: first at local level, the specific **benefits and challenges** and second, at global level, the complex connections by which they could be added or influence others in **superposition (less benefic) processes or consequences**.

This way, the most complicate and difficult problem of DD/ICT exponential evolutions management is **how to obtain, using both DD/ML/AI/ICT and HI resources, that refined knowledge which could provide optimal solutions to every relevant phase of those evolutions, i.e. how to save the Earth and humankind from irreversible consequences using the most powerful tools of science and technology (HI/AI)**.

Unfortunately, **this complicate question does not have a single (true) answer**, as we already approached before [16] [9], but people surely have to continuously try to find partially optimal solutions. More than these, the difficulty of this global problem is increasing with the level of complexity and with the speed of changes DD/ICT generate to IS/KBS.

It is sure that for this global problem, the engineers and other specialists involved in developing ICT systems, products and services have a prominent role (unfortunately not always decisive one). In fact, the fundamental problem of optimally refining the knowledge often starts at the technical designers, although **the opportunity of developing big projects should involve many levels and different areas specialists, in order to get a multicriterial optimization at global scale**.

As **opportunity** means also **what** knowledge must be refined, this problem is similar with the (university) conscious professor's dilemma, we all experience (desirably) every year at the beginning of school when we decide what to erase and what to introduce in our course, and generally *when a course, book or knowledge become obsolete*.

The final conclusion is, by our opinion, that, in our ever-changing DD days, it is almost impossible to precisely know when and how knowledge must be refined (a relative and approximative decision), but this does not mean that humankind have to ignore these challenges, on the contrary, we have to continuously look for getting as much as possible close to optimum, by the available updated data/information and resources, using AI/HI with desirable wisdom solutions.

References

- [1] Gordon Bell, Tony Hey, Alex Szalay, *Beyond the Data Deluge*, SCIENCE VOL 323 6 MARCH 2009
- [2]Javier D. Fernandez, Mario Arias, Miguel A. Martinez-Prieto, Claudio Gutierrez, *Management of Big Semantic Data*, November 2013, [https:// www.researchgate.net/ publication /259173878](https://www.researchgate.net/publication/259173878)

- [3]Victor Greu, *Using the information and communications technology data deluge from a semantic perspective of a dynamic challenge: What to learn and what to ignore?– (Part 1)*, Romanian Distribution Committee Magazine, Volume 10, Issue 3, Year 2019.
- [4]Shuai Zhang, Wenxi Zeng, I-Ling Yen, Farokh B. Bastani, *Semantically Enhanced Time Series Databases in IoT-Edge-Cloud Infrastructure*, 2019, CERN Open Days 2019 | 14–15 September, 2019, <https://arxiv.org> › pdf
- [5]***, *Artificial Intelligence and Machine Learning: Policy Paper*, April 2017, ISOC
<https://www.internetsociety.org/resources/doc/2017/artificial-intelligence-and-machine-learning-policy-paper/>
- [6]Victor Greu, *Extending information and communications technologies's impact on knowledge based society through artificial and collective intelligence – (Part 3)*, Romanian Distribution Committee Magazine, Volume 9, Issue 3, Year 2018.
- [7]Jorge Garcia, *Machine Learning and Cognitive Systems: The Next Evolution of Enterprise Intelligence*, 2014, <https://www.wired.com/insights/2014/07/machine-learning-cognitive-systems-next-evolution-enterprise-intelligence-part>.
- [8]Robert W. Lucky, *The Expiration Date of Knowledge*, IEEE Spectrum, March 2017.
- [9]Victor Greu, *The information and communications technology is driving artificial intelligence to leverage refined knowledge for the World sustainable development – (Part 2)*, Romanian Distribution Committee Magazine, Volume 10, Issue 1, Year 2019.
- [10]Maria Girone, *Tackling Tomorrow's Computing Challenges Today at CERN*, ISC 2018, Published in: Technology, Aug 1, 2018
- [11]Tony Hey, Anne Trefethen, *The Data Deluge: An e-Science Perspective*, <http://www.computing.surrey.ac.uk/courses/csm23/Papers/DataDeluge.pdf>], Wiley, 2003.
- [12]Bhupinder Kour, *The Rise of Machine Learning and AI is Improving Lives in 2018*, <https://www.smartdatacollective.com/rise-of-machine-learning-ai-improving-lives/>
- [13]Victor Greu et al, *Human and artificial intelligence driven incentive-operation model and algorithms for a multi-purpose integrated crowdsensing-crowdsourcing scalable system* - paper submitted to International Conference Communications 2018 (Politehnica University of Bucharest, Romania Military Technical Academy, IEEE Romania), June 2018.
- [14]Seth Earley, *There Is No AI Without IA*, 2016, IEEE IT Professional (Volume: 18, Issue:3, May-June 2016)
- [15]Niko Mohr, Holger Hürtgen, *Achieving business impact with data - A comprehensive perspective on the insights value chain*, Digital McKinsey 04.2018 Copyright © McKinsey & Company www.mckinsey.com
- [16]Victor Greu, *Searching the right tracks of new technologies in the earth race for a balance between progress and survival*, Romanian Distribution Committee Magazine, Volume 3, Issue1, Year 2012.
- [17]Mike Sirius, *How artificial intelligence and innovation will interact*, June 5, 2017, Idea Drop Ltd, <http://ideadrop.co/artificial-intelligence-innovation/>
- [18]Rohit Akiwatka, *Introduction to cognitive computing*, may 2019, <https://channels.theinnovationenterprise.com/articles/introduction-to-cognitive-computing>
- [19]Victor Greu, *The Exponential Development of the Information and Communications Technologies – A Complex Process Which is Generating Progress Knowledge from People to People*, Romanian Distribution Committee Magazine, Volume 4, Issue2, Year 2013
- [20]***, *Transforming data into knowledge*, Collective Learning group at the MIT Media Lab, 2017, <https://www.media.mit.edu/groups/collective-learning/overview/>.
- [21]Victor GREU, *Information and Communications Technologies are Learning from Nature's "Research" to Push the Performance Limits*, Romanian Distribution Committee Magazine, Volume 5, Issue 1, Year 2014
- [22]Mudassir Khan, *Big data analytics emerging trends, technology and innovations for the future business in the global market*, International Journal of Scientific Research and Review, Volume 8, Issue 2, 2019
- [23]Victor Greu, *Information and communications technologies go greener beyond IOT- behind is all the earth-Part1*, Romanian Distribution Committee Magazine, Volume 7, Issue 2, Year 2016