Information and Communications Technology is Merging Data Science and Advanced Artificial Intelligence Towards the Core of Knowledge Based Society

-Part 1-

Prof. Eng. Ph.D. Victor GREU

Abstract

The paper analyses the context of evolution of the Data Science (DS) concept, considering the importance of the mutual relations between the evolution of the technology/economy and generally of the society and the ways this is reflected in the people's minds, as knowledge, which is crucial as information and communication technology (ICT) became the main driving factor of the human society by the complex consequences of ICT services, products and applications when supporting the Information Society (IS) on the way towards the Knowledge Based Society (KBS). In fact, just the actual phase of the mutual relations between the evolution of the society and the knowledge across the individuals became such complex and potentially productive, so we consider it a strategic area for a timely evaluation and optimization, firstly for the obvious reason that the mechanisms behind these relations are building the foundation of the KBS. Although this phase is part of a process with a dynamic evolution in the last decades, the combination of the huge volume of data that are more and more generated on Earth (known as Data Deluge, Big Data etc.), with the naturally higher objectives of KBS, makes knowledge generation and its optimal use a real challenge for ICT specialists (and others too), which led to the DS concept and applications.

*The analysis started from the remarkable vision of Jim Gray on science evolution (**the fourth paradigm: the eScience**), just when the World and humankind are experiencing such complex and fast changes in the technology and science domains, facing unprecedented volume and diversity of available data which make difficult to estimate the future and the right strategy to approach it. We think, on the other hand, that, **in a complicate circle, ICT has the benefic potential, by continuously improving performances, to solve the complex problems which are associated with extracting information from the Data Deluge at Earth scale and eventually with the generation of knowledge that could be used by ICT too**. We also consider that **the fundamental and most difficult problem when evaluating these processes remain the content and profile of knowledge, which are time sensitive in a World where ICT exponentially generates changes in every human activity field**. This way we consider that solving such complex problems of evaluating knowledge generation and use has to be supported by the most performant ICT advances, where AI is by far the appropriate instrument for data/information high level processing. From these conclusions, the reasons for including DS and advanced AI in the ICT context for supporting the most critical core/part of KBS progress became clear, but this is just the tip of an iceberg we try to timely analyse step by step. The real challenges still begin just from here, because each of these 3 connected areas are very complex and their mutual relations are another complicate issue where knowledge is a common thread, as approaching them could be done only timely following all the impacted processes.*

We also concluded that evaluating knowledge generation and use is a matter of criteria that is essentially depending on human personality and generally on Earth ecosystem priorities, this way entering a complex circle with economical, social, philosophical and political main points of view. Just observing this (iceberg) circle of complicate interactions, we could understand why in our days it is more and more difficult to extract knowledge and especially to refine knowledge from the huge amount of generated data, if we want to continuously adapt/refine this knowledge to the optimal use and eventual effects for KBS.

Based on the fact that ACM (Association of Computing Machinery) recognized Jim Gray's vision on ICT by 1998 ACM Turing Award and in 1999 he expressed the new vision of ICT evolution as ***Cyberspace is a New World****, we considered that, while in the ICT exponential evolution any estimation is time sensitive, it is worth to analyse them by updating the implications of his vision due to the pillars he identified for this evolution.* ***Our opinion is that the essential feature of the ICT amazing evolution, which is manifest lately in each of its steps, is about the necessity to continuously re-invent (redefine) itself and learn from nature's millions of years research*** *and the practical value of Jim Gray's vision goes beyond its concrete pillars and features aimed to support the development of ICT, because it is* ***a sustainable set of ways of thinking about such complex and difficult approach like the ICT amazing evolution****.* ***Here we include a systemic way to spread knowledge and the need for it, across all individuals of KBS, beyond the usual approach of considering the ICT specialists, just to contribute and improve at least the interest for the necessary core of knowledge our paper aimed for KBS****. In addition, even if some of the goals could become sooner or later obsolete, we have to timely analyse the ICT evolution processes and imagine the best updated milestones, that will be surely better than having nothing ahead just for the reason that the challenges of prediction are overwhelming for such complex and complicate ecosystem at Earth scale.*

As we are in a spectacular development phase of this DS/AI/ICT/KBS context, its evolution naturally has to be further analysed, along with the appropriate features, consequences and learned lessons provided by ICT for all impact areas of the human activity and Earth ecosystem*.*

Keywords*: fourth paradigm of science, data deluge, big data, knowledge generation, data science, artificial intelligence, cyberspace, telepresence, automatic programmer, knowledge based society, eScience*

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

The only source of knowledge is experience -
Albert Einstein

1. Data science - The "fourth paradigm" of science

Watching movies with the first automobiles we can see how humans were scared, but along the centuries it seems that peoples became more habituated with such unprecedented technical wonders that massively get into their lives.

We could say that the way people perceive the surrounding picture is more and more influenced by technology, but even such largely shared opinion has to be suspected and analysed as a consequence of the evolution of the same technological context, which, among other things we actually see, is rapidly changing, not in centuries, not in decades, but day by day. From here, the sensation of a dizzy carousel is not far and the aim of understanding, where we are and the direction we are moving to, becomes elusive.

Still, even neglecting such, sometimes subjective, impressions, the problem of knowing **the importance of the mutual relations between the evolution of the technology/economy and generally of the society and the ways this is reflected in the people's minds, as knowledge, became crucial** as information and communication technology (ICT) became the main driving factor of the society.

On this way, we have repeatedly approached [9][3] the complex consequences of *ICT services, products and applications when supporting the Information Society (IS) on the way*

towards the Knowledge Based Society (KBS), but actualizing the extent of the mentioned importance and its concrete influence on future KBS remains crucial for the sustainable development of Earth.

In fact, just the actual phase of the mutual relations between the evolution of the society and the knowledge across the individuals became such complex and potentially productive, so we consider it a strategic area for a timely evaluation and optimization, firstly for the obvious reason that the mechanisms behind these relations are building the foundation of the KBS.

Although this phase is part of a process with a dynamic evolution in the last decades, the combination of the huge volume of data that are more and more generated on Earth (known as Data Deluge, Big Data etc.), with the naturally higher objectives of KBS, makes knowledge generation and its optimal use a real challenge for ICT specialists (and others too), which led to the Data Science (DS) concept and applications [4]:

<<This book presents the first broad look at the rapidly emerging field of data-intensive science, with the goal of influencing the worldwide scientific and computing research communities and inspiring the next generation of scientists. Increasingly, scientific breakthroughs will be powered by advanced computing capabilities that help researchers manipulate and explore massive datasets. The speed at which any given scientific discipline advances will depend on how well its researchers collaborate with one another, and with technologists, in areas of eScience such as databases, workflow management, visualization, and cloud-computing technologies. This collection of essays expands on the vision of pioneering computer scientist Jim Gray for a new, fourth paradigm of discovery based on data-intensive science and offers insights into how it can be fully realized….

Turing award winner Jim Gray imagined data science as a "fourth paradigm" of science (empirical, theoretical, computational and now data-driven) and asserted that "everything about science is changing because of the impact of information technology" and the data deluge.>>

Here we have to notice the remarkable vision of Jim Gray on science evolution (*the fourth paradigm: the eScience*), just when the World and humankind are experiencing such complex and fast changes in the technology and science domains, facing unprecedented volume and diversity of available data which make difficult to estimate the future and the right strategy to approach it, as it is also presented by [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.*”

On the other hand, **in a complicate circle, ICT has the benefic potential, by continuously improving performances, to solve the complex problems which are**

associated with extracting information from the Data Deluge at Earth scale and eventually with the generation of knowledge that could be used by ICT too.

Still, **the fundamental and most difficult problem when evaluating these processes remain the content and profile of knowledge, which are time sensitive in a World where ICT exponentially generates changes in every human activity field.**

This way we have just arrived to the point where it became clear that solving such complex problems of evaluating knowledge generation and use has to be supported by the most performant ICT advances, where artificial intelligence (AI) is by far the appropriate instrument for data/information high level processing.

In the same time, it is important to notice that evaluating knowledge generation and use is a matter of criteria that is essentially depending on human personality and generally on Earth ecosystem priorities, this way entering a complex circle with economical, social, philosophical and political main points of view.

Just observing this (iceberg) circle of complicate interactions, we could understand why in our days it is more and more difficult to extract knowledge and especially to refine knowledge from the huge amount of generated data, if we want to continuously adapt/refine this knowledge to the optimal use and eventual effects for KBS [10][11][12][19][13][17].

As a matter of fact, we have already mentioned [15] that approaching the human brain performance, AI could face just this sensitive feature of generation or extracting information from large amounts of data and more than these, of competing human brain for finding rules/solutions/knowledge even from smaller sets of data.

It is also worth to observe that knowledge always played a significant role for human progress, but actually it became overwhelming just due to the inherent capacity of knowledge to empower people to innovate, create and apply the powerful instruments of ICT in order to enable exponential changes in the whole Earth ecosystem, most (but not all) with benefic consequences for World sustainable progress.

Now it is clear that the premises of the paper title/aim include DS and advanced AI in the ICT context of supporting the most critical core/part of KBS progress, but this is just the tip of an iceberg we try to analyse step by step.

The real challenges still begin just from here, because each of these 3 connected areas are very complex and their mutual relations are another complicate issue where knowledge is a common thread, as approaching them could be done only timely following all the impacted processes.

As a result of advances in computing technology and data explosion, the complexity and the links between these 3 domains are also confirmed, in a comprehensive form, and then analysed focusing on DS, by [5]:

"*To solve big data problems in the era of big data and the data-driven paradigm, data science should incorporate the following factors: big data infrastructure, a big data analytics lifecycle, data management skills, and behavioural disciplines. Big data infrastructures include big data technologies such as Hadoop ecosystems, NoSQL databases, in-memory computing, as well as big data enabling technologies such as cloud computing."*

As it is indirectly above suggested, we also consider that DS is strongly linked with the education domain and this is then confirmed:

"*We proposed various approaches for data science education based on an extensive survey of current data science education programmes as well as domain knowledge in the field. To summarize, data science education should try to (1) teach CDO disciplines; (2) teach with the eight-step data analytics lifecycle in mind; (3) teach big data technologies and model-building techniques; (4) incorporate research methods in data analysis; (5) teach big data analytics as well as small data analytics; (6) provide students with real-world project experience; (7) collaborate with many departments; (8) collaborate with industry/ government for data, projects, resources, and practicums; and (9) actively use MOOCs= massive open online courses (MOOCs)*"

Unfortunately, but expected, the instruction in DS could not be covered simply by teaching the future chief data officers (CDO) by diverse courses:

"*We do not think that any single data science programme can adopt all of these approaches, and no data scientists could master all of these skills and knowledge that we have recommended. Each programme should focus on what they do best, and each data scientist should focus on what they do best. However, a certain level of broad coverage of topics is strongly recommended. We note the emergence of big data analytics with the usage of automated tools such as IBM Watson Analytics. Using automated tools or dashboards that use a black-box approach would be an important solution in training data scientists. However, the users of those tools should still be familiar with the methods implemented in the systems to choose a right method that fits the given data set and to interpret the outcomes properly. Those users should have critical thinking and reasoning ability to explore the solution space provided in the tool and to determine whether the tool can indeed provide a satisfactory outcome*"

More than these, we consider that, becoming a DS expert is a long process that should be timely updated, because having "*critical thinking and reasoning ability to explore the solution space*" is beyond of understanding the new tools, but in the same time understanding how Big Data appeared and what will mean in every area of applications, as it is also further indirectly mentioned:

"*How and why did the era of big data come about? Two of the major contributing factors to the emergence of the big data era include rapid advances in computing technologies and the resulting explosion of data; the former including hardware technologies such as CPU speeds and network bandwidths, as well as software technologies such as advent of distributed parallel processing frameworks (e.g. MapReduce and Hadoop); the latter including the increasing popularity of web-based software (e.g. search engines, social media networks, and e-commerce systems) as well as widespread usages of various sensors. These factors have collectively brought sudden explosion of data and contributed to the emergence of the big data era…How do people define big data? Gartner popularly defined it as 3Vs: 'high-volume, high-velocity, and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision-making' (Gartner, 2012b). Based on this definition, 4Vs were coined by adding the veracity dimension to the 3Vs, and 5Vs are also frequently mentioned by further adding the value dimension to the 4Vs.*"

Beyond Gartner recognized expertise in assessing the ICT evolutions, it is worth to notice the concrete/practical interpretations "*on experiences*" of the above V as:

"… (1) Volume means the size of data that scales to terabytes, petabytes, or even more. We view volume as a technology solution as we can easily buy those technologies. (2) Velocity

means the speed of creating/processing/analysing/storing data. We view velocity as a semi-technology solution, as we can buy some solutions, but we still need to develop creative software to handle them. (3) Variety means different data types, sources, and modes to handle. We view variety as a software solution, as there are still many remaining challenging software issues that need to be addressed. (4) Veracity means quality, reliability, and uncertainty in data. We view veracity as a challenging research dimension, as it is an area that still needs to be more thoroughly researched, especially on the impacts of veracity to data integration and analytics. (5) Value means the discovery of actionable knowledge, high return on investment, increased relevancy to customers or products, or innovations in business operations/processes. We consider value as the most important V in the big data era. Without extracting value from big data, big data projects would not be meaningful. While we view variety and veracity as challenging dimensions, value is by far the most challenging dimension. If we are able to address these challenges and extract value from big data, then big data projects will give us opportunities for innovative solutions and chances to make an impact on technology, society, and business. We are beginning to see profound impacts of big data in every aspect of our lives and society"

This way (*profound impacts of big data in every aspect of our lives and society*), our paper point is clearly confirmed and the **reasons, for further and deep analyses of the processes that link DS with advanced AI toward generating knowledge to support the core of KBS, are revealed as influencing all the levels and directions of developing the World** (*an impact on technology, society, and business*).

2. Cyberspace as a *New World*

The evolution of terms which are associated with ICT/KBS context and the necessary strategies for a sustainable development are also issues that are worth to be analysed.

For the same context, ACM (Association of Computing Machinery) recognized Jim Gray's vision on ICT by 1998 ACM Turing Award and we also consider important to recall that in 1999 he expressed the new vision of ICT evolution as ***Cyberspace is a New World*** [1].

Because in the ICT exponential evolution any estimation and even term is time sensitive, it is worth to notice that the updated analysis and vision of Jim Gray on ICT context still remain remarkable and with actual importance, due to the pillars he identified for this evolution [1]:

"…long-range research has societal benefits, both in creating new ideas and in training people who can make even better ideas and who can turn those ideas into products. The education component is why much of the research should be done in a university setting. This argues for government support of long-term university research…"

It is important to observe the same confirmation of the fundamental "circle" (where knowledge is generated and used) we have above pointed, as "*training people who can make even better ideas and who can turn those ideas into products*".

In addition, Jim Gray also confirms the strategic role of education for the KBS and consequently argued for *government support of long-term university research*.

As we have repeatedly mentioned [11][18] and pointed above, when experiencing such complex and fast changes in the technology and science domains it is difficult to estimate the

future and the right strategy to approach ICT/KBS, Earth ecosystem and humankind evolution, i.e., what Jim Gray did, just because *Exponential Growth Means Constant Radical Change* [1]:

"*Exponential growth has been driving the information industry for the last 100 years. Moore's law predicts a doubling every 18 months. This means that in the next 18 months there will be as much new storage as all storage ever built, as much new processing as all the processors ever built… In 1995, George Glider predicted that deployed bandwidth would triple every year, meaning that it doubles every 8 months. So far his prediction has been pessimistic: deployed bandwidth seems to be growing faster than that!*"

More than these, as we also mentioned [6], ICT is in a continuous struggle to maintain this growing, although it is harder and harder to succeed that, even by inventing new radical advances:

"*Exponential growth cannot go on forever. E. coli (bacteria in your stomach) double every 20 minutes. Eventually something happens to limit growth. But, for the last 100 years, the information industry has managed to sustain this doubling by inventing its way around each successive barrier. Indeed, progress seems to be accelerating … Some argue that this acceleration will continue, while others argue that it may stop soon – certainly if we stop innovating it will stop tomorrow. These rapid technology doublings mean that information technology must constantly redefine itself: many things that were impossibly hard ten years ago, are now relatively easy. Tradeoffs are different now, and they will be very different in ten years.*"

Our opinion is that here the essential feature of the ICT amazing evolution is excellently pointed as: "*information technology must constantly* ***redefine itself****: many things that were impossibly hard ten years ago, are now relatively easy*". Generally, we also mentioned that the ICT exponential evolution could only based on ***the necessity to continuously re-invent itself and learn from nature's millions years research***[18].

Consequently, Jim Gray expressed the ICT evolution and premises (*Cyberspace is a New World*) as the place of a new revolution*:*

"One way to think of the Information Technology revolution is to think of cyberspace as a new continent -- equivalent to discovery of the Americas 500 years ago. Cyberspace is transforming the old world with new goods and services. It is changing the way we learn, work, and play"

It is worth to notice that, apparently in contradiction with the mentioned "*rapid technology doublings*", some old references in the same area of predicting ICT evolution are important and worth to be considered:

*<<… recall Alan Turing's famous "Computing Machinery and Intelligence" paper published in 1950****Error! Reference source not found.****. Turing argued that in 50 years, computers would be intelligent. This was a very radical idea at that time. Turing's actual text on this matter is worth re-reading. What he said was:*

"I believe that in about fifty years' time it will be possible, to programme computers, with a storage capacity of about 10^9*, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning. The original question, "Can machines think?" I believe to be too meaningless to deserve discussion. Nevertheless, I believe*

that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.">>

This incredible evolution and prediction are explained by identifying another of the mentioned pillars of ICT/KBS, we called **circle** evolution, as "**long-term research is a social good**" [1]:

"*One reason for this is that long-term research is a social good, not necessarily a benefit to the company. AT&T invented the transistor, UNIX, and the C and C++ languages. Xerox invented Ethernet, bitmap printing, iconic interfaces, and WYSIWYG editing. Other companies like Intel, Sun, 3Com, HP, Apple, and Microsoft got the main commercial benefits from this research. Society got much better products and services -- that is why the research is a public good*."

By Jim Gray's vision, the sustainable evolution of ICT should include some other reliable pillars which were expressed as:

"8. *TelePresence: Simulate being some other place retrospectively as an observer (TeleOberserver): hear and see as well as actually being there, and as well as a participant, and simulate being some other place as a participant (TelePresent): interacting with others and with the environment as though you are actually there.*

...9. Trouble-Free Systems: Build a system used by millions of people each day and yet administered and managed by a single part-time person

...10. Secure System: Assure that the system of problem 9 only services authorized users, service cannot be denied by unauthorized users, and information cannot be stolen (and prove it.)

... 11. AlwaysUp: Assure that the system is unavailable for less than one second per hundred years -- 8 9's of availability (and prove it.)"

Perhaps *telepresence* could be considered not very relevant on long term, but lets just make an exercise of imagination, beyond the obvious benefits of efficiently sharing experience and real time collaboration in the research and technology areas: Which was our opinion on telepresence before Covid 19 pandemic and which is now?

Another opinion test should come if imagining the possible consequences of climate (agressive) changes, Earth resources fading or economical/social crises/unbalances at planetary scale.

Although the features 9, 10 and 11 could be seen similarly, as mainly concerning the safety areas, we consider that finally the net reliability (*unavailable for less than one second per hundred years -- 8 9's of availability*) is a crucial feature, considering the strategic impact of ICT on all activity fields, but mainly on the critical infrastructures.

Here, the number of 9's is very significant by its evolution, as it is also further detailed:

"*We have gone from 90% availability in the 1950s to 99.99% availability today for well managed systems. Web uses experience about 99% availability due to the fragile nature of the web, its protocols, and the current emphasis on time-to-market.*

Nonetheless, we have added three 9s in 45 years, or about 15 years per order-of-magnitude improvement in availability. We should aim for five more 9s: an expectation of one second outage in a century. This is an extreme goal, but it seems achievable if hardware is very cheap and bandwidth is very high. One can replicate the services in many places, use transactions

to manage the data consistency, use design diversity to avoid common mode failures, and quickly repair nodes when they fail. Again, this is not something you will be able to test: so achieving this goal will require careful analysis and proof."

We appreciate that, among other pillars, a special relevance has and always will have the *Automatic Programmer*, i.e. one of the most advanced skills/features of AI that could compete with the human brain in the future:

"…12. Automatic Programmer: Devise a specification language or user interface that:
(a) makes it easy for people to express designs (1,000x easier),
(b) computers can compile, and
(c) can describe all applications (is complete)."

The system should reason about application, asking questions about exception cases and incomplete specification. But it should not be onerous to use."

Although in such fast changing domain like ICT it is difficult to keep the coarse for a long time even with a good strategic plan, in this case the Jim Gray's remarkable vision included some featureas (*What Makes a Good Long Range Research Goal?*) that could just attenuate the chances that vision content will become obsolete:

"*Understandable: The goal should be simple to state. A sentence, or at most a paragraph should suffice to explain the goal to intelligent people. Having a clear statement helps recruit colleagues and support. It is also great to be able to tell your friends and family what you actually do.*

Challenging: It should not be obvious how to achieve the goal. Indeed, often the goal has been around for a long time. Most of the goals I am going to describe have been explicit or implicit goals for many years. Often, there is a camp who believe the goal is impossible.

Useful: If the goal is achieved, the resulting system should be clearly useful to many people -- I do not mean just computer scientists, I mean people at large.

Testable: Solutions to the goal should have a simple test so that one can measure progress and one can tell when the goal is achieved.

Incremental: It is very desirable that the goal has intermediate milestones so that progress can be measured along the way. These small steps are what keep the researchers going."

We consider that the importance of this strategic vision goes beyond its concrete pillars and features aimed to support the development of ICT, because it is a sustainable set of ways of thinking about such complex and difficult approach like the ICT amazing evolution.

Among these strategic ideas we would remark that "*to be able to tell your friends and family what you actually do*" is not a simply advice about DS/ICT image, but a systemic way to spread knowledge and the need for it, across all individuals of KBS (*I mean people at large*), just to contribute and improve at least the interest for the necessary core of knowledge.

In spite of the complexity and difficulty of the approached research fields, the features of simplicity, that are very clearly suggested, at goal, testing or progress steps, are very important for the realistic approach and sustainable development of DS/ICT/KBS context (Cyberspace *New World*).

In addition, our opinion is that, even if some of the goals could become sooner or later obsolete, we have to timely analyse the ICT evolution processes and imagine the best updated milestones, that will be surely better than having nothing ahead just for the reason that the

challenges of prediction are overwhelming for such complex and complicate ecosystem at Earth scale.

In fact, we could observe that we are in a spectacular development phase of this ICT/KBS context, at Earth scale, that, continuing Jim Gray's vision, is largely recognized and characterized, in the last decade, by terms like ***Cyberinfrastructure*** [7][8][14][16][5], which naturally have to be further analysed, along with the appropriate features, consequences and learned lessons provided by ICT for all impact areas of the human activity and Earth ecosystem.

3. Conclusions

We consider that the importance of the mutual relations between the evolution of the technology/economy and generally of the society and the ways this is reflected in the people's minds, as knowledge, became crucial as information and communication technology (ICT) became the main driving factor of the human society and also we repeatedly mentioned [9][3] the complex consequences of *ICT services, products and applications when supporting the Information Society (IS) on the way towards the Knowledge Based Society (KBS)*, but actualizing the extent of the mentioned importance and its concrete influence on future KBS remain relevant for the sustainable development of Earth.

In fact, just the actual phase of the mutual relations between the evolution of the society and the knowledge across the individuals became such complex and potentially productive, so we consider it a strategic area for a timely evaluation and optimization, firstly for the obvious reason that the mechanisms behind these relations are building the foundation of the KBS. Although this phase is part of a process with a dynamic evolution in the last decades, the combination of the huge volume of data that are more and more generated on Earth (known as Data Deluge, Big Data etc.), with the naturally higher objectives of KBS, makes knowledge generation and its optimal use a real challenge for ICT specialists (and others too), which led to the Data Science (DS) concept and applications.

Our analysis started from the remarkable vision of Jim Gray on science evolution (*the fourth paradigm: the eScience*), just when the World and humankind are experiencing such complex and fast changes in the technology and science domains, facing unprecedented volume and diversity of available data which make difficult to estimate the future and the right strategy to approach it.

We concluded, on the other hand, that, **in a complicate circle, ICT has the benefic potential, by continuously improving performances, to solve the complex problems which are associated with extracting information from the Data Deluge at Earth scale and eventually with the generation of knowledge that could be used by ICT too**.

We also consider that **the fundamental and most difficult problem when evaluating these processes remain the content and profile of knowledge, which are time sensitive in a World where ICT exponentially generates changes in every human activity field**. This way we have just pointed that solving such complex problems of evaluating knowledge generation and use has to be supported by the most performant ICT advances, where AI is by far the appropriate instrument for data/information high level processing

From these conclusions, the reasons for including DS and advanced AI in the ICT context for supporting the most critical core/part of KBS progress became clear, but this is just

the tip of an iceberg we try to timely analyse step by step. The real challenges still begin just from here, because each of these 3 connected areas are very complex and their mutual relations are another complicate issue where knowledge is a common thread, as approaching them could be done only timely following all the impacted processes.

We also concluded that evaluating knowledge generation and use is a matter of criteria that is essentially depending on human personality and generally on Earth ecosystem priorities, this way entering a complex circle with economical, social, philosophical and political main points of view. Just observing this (iceberg) circle of complicate interactions, we could understand why in our days it is more and more difficult to extract knowledge and especially to refine knowledge from the huge amount of generated data, if we want to continuously adapt/refine this knowledge to the optimal use and eventual effects for KBS.

As ACM (Association of Computing Machinery) recognized Jim Gray's vision on ICT by 1998 ACM Turing Award and in 1999 he expressed the new vision of ICT evolution as *Cyberspace is a New World*, we considered that, while in the ICT exponential evolution any estimation and even term is time sensitive, it is worth to analyse them by updating the implications of his vision due to the pillars he identified for this evolution.

Our opinion is that the essential feature of the ICT amazing evolution, which is manifest lately in each of its steps, is about the necessity to continuously re-invent (*redefine*) itself and learn from nature's millions of years research. The practical value of Jim Gray's vision goes beyond its concrete pillars and features aimed to support the development of ICT, because it is **a sustainable set of ways of thinking about such complex and difficult approach like the ICT amazing evolution**. **Here we include a systemic way to spread knowledge and the need for it, across all individuals of KBS, beyond the usual approach of considering the ICT specialists, just to contribute and improve at least the interest for the necessary core of knowledge our paper aimed for KBS**. In addition, even if some of the goals could become sooner or later obsolete, we have to timely analyse the ICT evolution processes and imagine the best updated milestones, that will be surely better than having nothing ahead just for the reason that the challenges of prediction are overwhelming for such complex and complicate ecosystem at Earth scale.

Because we are in a spectacular development phase of this DS/AI/ICT/KBS context, its evolution naturally has to be further analysed, along with the appropriate features, consequences and learned lessons provided by ICT for all impact areas of the human activity and Earth ecosystem.

REFERENCES

[1]Jim Gray, *What Next?A Dozen Information-Technology Research Goals*, Microsoft Research, June 1999, Technical Report MS-TR-99-50

[2]Javier D. Fernandez, Mario Arias, Miguel A. Martınez-Prieto, Claudio Gutierrez, *Management of Big Semantic Data*, November 2013, https:// www.researchgate.net/ publication /259173878

[3]Florin Enache, Victor Greu, Petrică Ciotîrnae, Florin Popescu, *Model and Algorithms for Optimizing a Human Computing System Oriented to Knowledge Extraction by Use of Crowdsourcing*, 2020, 13th International Conference on Communications (COMM), (Politehnica University of Bucharest, Military Technical Academy, IEEE Romania), (COMM 2020 is covered in IEEE Explore Database and ISI Web of Science in the Conference Proceedings Citation Index)

[4]Tony Hey, *The Fourth Paradigm: Data-intensive Scientific Discovery*, Microsoft Research, 2009 - Science
[5]Il-Yeol Song and Yongjun Zhu, *Big data and data science: What should we teach?*, Expert Systems, August 2016, Vol. 33, No. 4, https://www.researchgate.net/publication/282692841
[6]Victor Greu, *Searching the right tracks of new technologies in the earth race for a balance between progress and survival*, **Romanian Distribution Committee** (affiliated to the "International Association of the Distributive Trade"-scientific association – A.I.D.A. Brussels) **Magazine** (international; electronic; covered in RePEc International Data Base), Volume 3, Issue1, Year 2012.
[7]Craig A. Stewart et all, *What is Cyberinfrastructure?*, SIGUCCS'10, October 24–27, 2010, Norfolk, Virginia, USA., https://www.researchgate.net/publication/49471518
[8]Daniel E. Atkins et al, *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*, 2003, https://www.nsf.gov/cise/sci/reports/atkins.pdf
[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 (affiliated to the "International Association of the Distributive Trade"-scientific association – A.I.D.A. Brussels) Magazine(international; electronic; covered in RePEc International Data Base), Volume 10, Issue 1, Year 2019.
[10]Robert W. Lucky, "*The expiration date of knowledge [Opinion]*", IEEE Spectrum, vol. 56, no. 09, pp. 21-21, Sept. 2019
[11]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 2)*, Romanian Distribution Committee (affiliated to the "International Association of the Distributive Trade"-scientific association – A.I.D.A. Brussels) Magazine(international; electronic; covered in RePEc International Data Base)
, Volume 10, Issue 4, Year 2019.
[12]Gerbrand Tholen, *The problem with a knowledge-based society*, November 26th 2017, https://blog.oup.com/authors/gerbrand-tholen/
[13]Victor Greu, *Extending information and communications technologies' impact on knowledge based society through artificial and collective intelligence –(Part 3)*, Romanian Distribution Committee (affiliated to the "International Association of the Distributive Trade"-scientific association – A.I.D.A. Brussels) Magazine(international; electronic; covered in RePEc International Data Base), Volume 9, Issue 3, Year 2018.
[14]E. S. Vorm, *Computer-Centered Humans: Why Human-AI Interaction Research Will Be Critical to Successful AI Integration in the DoD*, IEEE Intelligent Systems (Volume: 35, Issue: 4, July-Aug. 1 2020),https://ieeexplore.ieee.org/document/9179109/authors#authors
[15]Victor Greu et all, *Human and artificial intelligence driven incentive-operation model and algorithms for a multi-purpose integrated crowdsensing-crowdsourcing scalable system*, Proceedings of International Conference Communications 2018, (Politehnica University of Bucharest, Military Technical Academy, IEEE Romania), June 2018(COMM 2018 is covered in IEEE Explore Database and ISI Web of Science in the Conference Proceedings Citation Index).
[16]Fran Berman, *Current Working Definitions Of Cyberinfrastructure*, 2005, https://www.researchgate.net/publication/49471518_What_is_Cyberinfrastructure/link/0912f51085bdceaf0a000000/download
[17]Tony Hey, Anne Trefethen, *The Data Deluge: An e-Science Perspective*, http://www.computing.surrey.ac.uk/courses/csm23/Papers/DataDeluge.pdf], Wiley, 2003.
[18]Victor GREU, *Information and Communications Technologies are Learning from Nature's "Research" to Push the Performance Limits*, Romanian Distribution Committee (affiliated to the "International Association of the Distributive Trade"-scientific association – A.I.D.A. Brussels) Magazine(international; electronic; covered in RePEc International Data Base), Volume 5, Issue 1, Year 2014
[19]Gordon Bell, Tony Hey, Alex Szalay, *Beyond the Data Deluge*, Science Vol 323 6 March 2009