On Big Science: A Survey

Kemal A. Delic
Hewlett-Packard Co.
Enterprise Services
kemal.delic@hp.com

Jeff A. Riley
Hewlett-Packard Co.
Enterprise Group
jeff.riley@hp.com

ABSTRACT
Scientific quest is driven by eternal human curiosity and the need to understand and explain unknown worlds. It is an endless process of extending current human knowledge with new insights, improved theories, and advances via discoveries and singular inflection points. Big Science denotes long lasting, high risk and extremely expensive projects addressing the big challenges and hard issues. Opinions are divided about the Big Science approach to scientific advances, and this short paper setts the ground for a discussion about the merits and perils of such an approach. We conclude with our belief that the changed technological circumstances, recent scientific achievements, and economic needs will likely lead to a new rise of experimental sciences. We believe that a combination of AI-inspired methods, digesting unprecedented volumes of data and exploiting the limitless capacities of computing clouds, will create new kinds of scientific instruments, tools, and collaborative environments, advancing theoretical insights via experimental proofs.

CCS Concepts
• General and reference➞Surveys and overviews • Computing methodologies➞Artificial intelligence • Computer systems organization➞Architectures➞Distributed architectures➞ Cloud computing • Information systems organization➞ Information systems applications➞Data mining

Keywords
Big Science; Artificial Intelligence; Cloud Computing; Big Data

1. INTRODUCTION
Human curiosity about nature has driven scientific explorations over thousands of years. Those explorations have passed from a primitive interpretation of the world, via constantly improving sophistication of theories and experimentations, until the moment when critical mass has been reached, enabling great scientific discoveries. 'Big Science' refers to large, ambitious projects, aimed at resolving long-standing mysteries/uncertainties or engaging large scientific communities against some big, eternal challenges. We argue here that the co-occurrence and corroboration of Big Data, Internet of Things and Artificial Intelligence may create fertile ground for Big Science. We also postulate that this approach could lead to a revival of Big Experimental Sciences discussed since the 1960s.

We start with a very short history of advances through science, then decompose Big Science into three components which, we believe, should play major roles in the future. Then we describe three remarkable Big Science projects, and conclude with our belief that all this may result in the rise of Big Experimental Sciences.

ACM 978-1-4503-3721-2/15/10…$15.00

DOI: http://dx.doi.org/10.1145/2809563.2809622

2. BRIEF HISTORY OF THE SCIENCE

Since the appearance of the first humans on the Earth, surrounding nature, curios viewed in the night sky, changing weather phenomena and observed cyclical changes resulted in the first, naïve interpretations of observations passed orally down through the generations until writing was invented. We could call it 'primitive science'. Civilizations in Egypt and Mesopotamia advanced human knowledge and skills, and captured it in clay tablets for the posteriori and preservation. Early Greece gave us several important natural philosophers and mathematicians, laying the ground for later advances. Chinese science produced many innovations (e.g. paper money, silk, gunpowder) which were not well known in the Western world. The rising influence of Islamic science was based on their translation of Greek sciences, which they enriched and passed to the Renaissance age. Several distinctive individuals (e.g. Galileo, Da Vinci, Newton) marked the golden age of medieval sciences. Entering the industrial period, we have seen practical developments of various technologies, advances in mathematics, medicine and natural sciences, typically brought forward by brilliant individuals. Modern science probably begins in 20th century, with important technological advances accelerated by two world wars and the rising challenges of a quickly rising world population [1]. To set the stage, we depict the role of science in the modern, contemporary world (Fig. 1).

Society has a strong interest to cultivate, grow and enrich scientific research (1), which will in turn create the basis for a wide variety of technologies (2), which will then be employed in the economy (3) to create jobs, wealth and prosperity in society (4). It would be the question of established policy or political decisions of government/international bodies about what percentage of investments are to be re-invested into scientific research. It is nearly accepted truth that prosperous societies benefit most from such re-investments. This is, of course, an over-simplification for the sake of arguments supporting Big Science adventures. The real world circumstances are infinitely more complex. In its simplest possible form, funding decisions are seen as investment choices between one big and several small investments. To discuss perspectives of potential, future Big Science projects, we discuss briefly three principal ingredients which we believe will be important for the future.

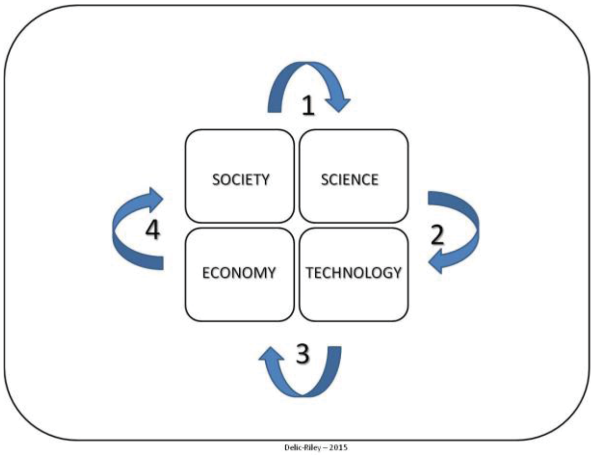

Figure 1. Simplistic view of contemporary world

3. BIG SCIENCE COMPONENTS

The beginning of the 21st century was marked by a steep rise of digitalization, where for the first time in history the volume of digital data surpassed the volume of analog data. This steep rise continues today – as web growth, exhaustive infrastructure monitoring, and sensor developments create an unprecedented tsunami of **Big Data**. To cope with it, **Computing Clouds** are ideal fabrics to store, process and move this data into appropriate form factors and deliver it for ultimate consumption. In addition, new types of algorithms have been born over the past 60 years of developments in **Artificial Intelligence**, being able to emulate human intelligence, cognition and behavior.

3.1 Big Data

Recent advances in the field of Information Technologies (IT) leading to massive consumerization of IT, followed by a rapid drop in IT prices, have resulted in a world of 2 billion mobile, nearly-always connected devices, creating and consuming an avalanche of data (**IoT – Internet of Things**). In addition, we see an increasing deployment of a variety of sensors on humans, devices and infrastructural objects. Huge instruments are created, generating petabytes of streaming data, and the sky is explored with radio telescopes producing immense quantities of high-resolution visual and telemetry data. Beyond the practical problem of what to do with all this data – called Big Data challenge – there is a new profession of **Data Scientist**, aiming to deal with data in a different, innovative manner. To deal with the scale, volume and dynamics of big data, (data) scientists will need big resources.

3.2 Computing Clouds

Computing Clouds are huge aggregates of computing, storage and networking resources, used in a pay-for-service manner. Cloud computing has its roots in the time-share systems of the 1950s, allowing many users to share the computing power of large mainframes. This practice developed further with Virtual Machines in the 1970s. The development of ARPANET in the 1960s and 70s as part of the quest for a networking solution able to survive nuclear attacks, was the first step towards Licklider's early vision "*for everyone on the globe to be interconnected and accessing programs and data at any site, from anywhere*" [2]. The advent of Virtual Private Networks in the 1990s allowed large companies to expand their private networks across the globe. The development of web servers and clients at CERN in the 1990s, embodying the HTTP communication protocol and HTML page description language, created the Web as we know it today. With personal computers making computing more affordable and accessible to the general population in the 1980s, and smart handheld devices such as tablets and smart phones becoming ubiquitous in the 2000s, billions of devices around the world are now able to connect to any one of many computing clouds developed to offer compute, storage and network resources as a service.

It has quickly become obvious that this may change entirely the scale of scientific endeavors [3], and open-up unexpected avenues of future research [4].

3.3 Artificial Intelligence

Once we have massive data sets loaded onto computing clouds, they would need to be processed with intricate algorithms which will (somehow) emulate human interpretation, analysis and synthesis of the large data sets. Thus, they will emulate perceptual and cognitive capabilities of humans at a scale and speed that no human group can match. Already large amounts of data created/consumed on social networks are digested and synthesized into 'sentiment analyses', 'influence maps' and 'crowd perception' artifacts – as telling examples of the results of the deployment of clever, parallel algorithms on large-scale infrastructure.

We can easily envisage a wide variety of investigations of collections of scientific data to digest and produce results similar to human-produced results. AI will not replace scientists, but it does, and will continue to, augment their capabilities to deal with huge data volumes and rates, and accelerate results of better quality.

This esoteric research into the nature of human intelligence [5] during the last 60 years has passed the three principal epochs: **embryonic** intelligence **embedded** intelligence and **embodied** intelligence. [6].

For many years, AI was a field of inspiration for software developers and algorithm theorists. It will certainly play an even bigger role in the future, as we see current commercial systems deploying AI algorithms at a scale never tried before. We expect that soon those systems will be deployed on large streams of scientific data – as is the case with CERN using AI algorithms to explore and digest 140 PB of raw data [7].

4. SOME CASES OF BIG SCIENCE PROJECTS

We will now briefly describe some well-known big science projects and conclude with a few forward-looking thoughts. The projects range from space exploration via telescopes and landing of several automatic laboratories onto a wandering comet, to the subatomic world of quantum computing and nano-technologies. Particular recent focus is on human brain research, cancer diseases and genomics. Alternative energy sources based on nuclear fusion instead of fission are highly sought after. For the sake of our arguments we will give some data points for a few selected projects.

The Large Hadron Collider (**LHC**) is the largest and most complex experimental facility ever made. The LHC cost 9B$ during 10 years of construction, involving more than 10,000 scientists from over 100 countries. ATLAS is the system capturing collision traces and, up until now, it has generated around 140 PB of raw data. This huge instrument has proved the existence of a theoretically posited sub-atomic particle, and has produced a Nobel Prize laureate in Physics. More exciting, promising and exploitable results will be delivered in the future [8].

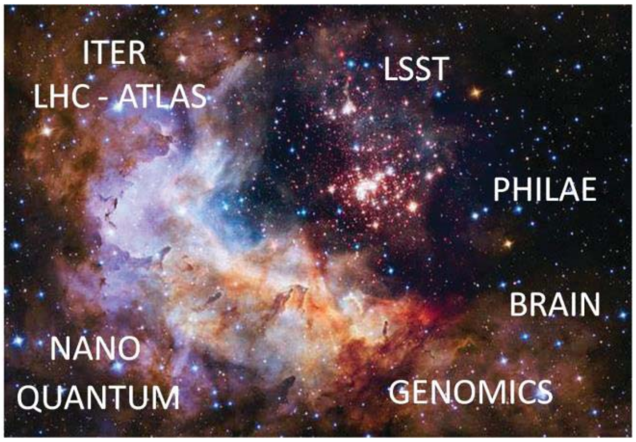

Figure 2. Constellation of the Big Data projects

NASA's James Webb Space Telescope (**JWST**) is currently running Big Science project experiencing budgeting and management issues which are currently fixed - as a good example of NASA project governance improvements [9].

The **ITER** project (International Thermonuclear Experimental Reactor) is run by an international consortium of seven members, and has been running for ten years, costing 4.4 B$. Project **ITER** aims to provide alternative energy sources by developing a nuclear fusion reactor. More than 2000 people are working on the project currently [10] – it is an illustration of long-running political will (since 1985) to address problems of energy generation.

The European Space Agency (ESA) worked for 15-20 years on the **Rosetta - PHILAE** project – a mission to land a vehicle on a comet after 11 years of travel through space. The landing module Philae has several on-board labs through which it is able to execute scientific experiments, gather results and send data, together with high-resolution images, back to Earth. It is an historical event - nothing similar has been done before - and the hope is to better understand and explain the formation of the solar system and birth of life on earth [11].

The Large Synoptic Survey Telescope (**LSST**), scheduled to go live in 2020, will capture high-resolution image of the sky over a period of 10 years, storing 100+ Petabytes for post-analysis - this is yet another example of a unique, expensive scientific instrument creating a huge repository of visual data [12].

The European Union is funding a 10-year, 1.1B$ program to explore carbon **nanostructures** which may change the entire field of material sciences and have a huge commercial impact [13].

Very recently, both the US and the EU have launched 10-year, 1+B$ projects in **BRAIN** research. Right now there is a discussion about Human Brain Project (**HBP**), its stated scientific goals and the adequacy of its management structure. [14].

GENOMICS – the Human Genome Project (**HGP**) remains the world's largest collaborative biological project, and spent nearly 3B$ over 15 years in order to analyze more than 100,000 human genomic fragments[15]. Following recent developments and cost reductions, insights from this project will likely have a very big commercial impact. However, we should not forget the largely disregarded issue of privacy protection – as we could easily imagine the consequences of early prediction of possible disease on the private and professional lives of individuals. Imagine the impact of medical records combined with financial records used in professional circumstances. Thus, we postulate that this domain should be highly regulated and constantly monitored, and note that the potential for innovation here is very high.

A high level overview of these projects will indicate the same characteristic repetitive pattern – it is always a **big data problem**, requiring **big resources**, **long-lasting budget**, important **reserve of nerves** and **brave persistence.** Also, they are typically restructured in the middle of execution; objectives are re-adjusted, and more often than not promises are fulfilled. Discussing all the impacts and benefits of Big Science projects is too complex to be treated in this short paper. However, debate about it is still on-going, conclusions are not definitive, and confronting camps are created in government and policy circles advocating Big versus Small advantages. We believe that the final verdict cannot be given in the current circumstances.

5. OUTLOOK – CONCLUSION

The long history of science has seen modest beginnings, but always accelerating developments in which some brave, rare and brilliant individuals advanced the state of human knowledge [16]. Today, hundreds of thousands of scientists are working in a wide variety of research fields, bringing new knowledge daily and developing useful technologies for commercial purposes. In the current circumstances, the scale, scope and speed of scientific research has reached unprecedented values. We postulated that the three ingredients may represent good ground for the next wave of Big Science projects.

We strongly believe that Big Science is creating a large number of commercial technologies as payoff for equally large initial investments: we see examples of this from investments in space exploration, resulting in the development of new materials, new communication technologies, automatic labs, new devices etc. We believe this indicates it is better to invest in a small number of Big Science projects rather than many fragmented small projects.

Not a very long time ago (in 2008), the editor-in-chief of Wired magazine announced **The Petabyte Age** and **The End of Theory**, creating vivid conversations and hinting at a new experimental age. Seven years later, we see that he was largely right, and new commercial enterprises are thriving on data tsunamis, and new scientific instruments and methods will be born in all scientific fields. In the long historical perspective, the development of instruments (e.g. telescope, microscope, etc.) has always triggered a new kind of sciences, created new insights and corroborated theoretical assumptions. In the same vein, we should not disregard the great importance and contribution of old-fashioned, advanced mathematics which will be augmented with big data collections coming from experimentation [17].

Our key hypothesis (to be proven) is that an intricate combination of AI-inspired methods, digesting unprecedented volumes of data, and exploiting the limitless capacities of computing clouds will create new kinds of instruments, tools and collaborative environments, advancing theoretical insights via experimental proofs.

We are not suggesting that the combination of three important fields will create a new science, but we envisage a positive context in which some new instruments and tools, as well as distributed, collaborative environments (cf. the rise of social networks), will be created based on advances in those three fields. Ultimately, those instruments will not replace scientists, but will largely improve, enhance and augment scientists' capabilities. Thus, science (the body of knowledge) as the advanced knowledge contributing to the evolution of human kind, will be better supported by augmenting the capabilities of science practitioners.

One would then try to understand what is triggering Big Science Projects. This may shed some light onto motivational factors, project execution context, and provide a reality check on impacts and results.

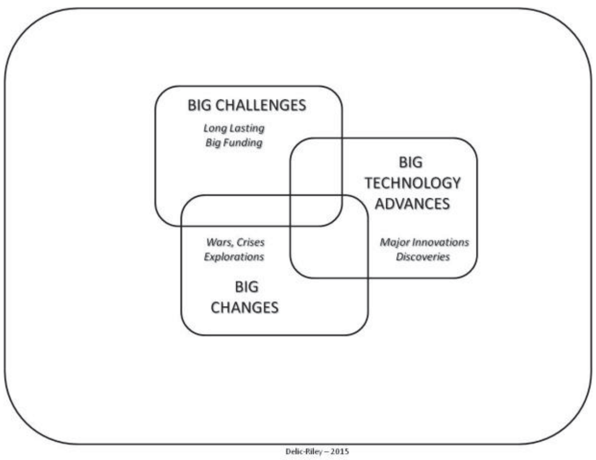

Figure 3. Typical Drivers of Big Science projects

Looking back into past Big Science project triggers, we can group them into three specific groups (Fig. 3). Unfortunate events such as wars and crises which change landscapes and power balances dramatically are often big technological accelerators. Major innovations and discoveries can move entire industries and society segments forward. Finally, the most frequent trigger of Big Science projects is big, government, international funds posing huge challenges [18]. Ultimately, it is quite often that an intricate mix of all three factors is motivation to fund those risky adventures. In conclusion, we see all these developments leading to the rebirth of **The New Experimental Sciences**, based on the three substantial components described in section 3. We also envisage that a new combination of known methods will create innovative new methods, stipulated by deployment of new technologies: such has been the case in past innovation and discovery.

6. ACKNOWLEDGMENTS

We thank Dr Ulrich Reimer for his insightful comments, and anonymous reviewers for providing useful comments and suggestions.

This short poster/position paper addresses a huge domain and it was conceived as a short script for a live, round-table conversation which will likely be held at iKNOW'16. As such, it should be not treated as a research or scientific paper, but a large canvas landscaping exercise.

7. REFERENCES

[1] Colin Ronan, 1983. *The Cambridge Illustrated History of the World's Science,* Cambridge University Press. 1983.

[2] J.C. R. Licklider, 1963. "Topics for discussion at the forthcoming meeting, memorandum for: members and affiliates of the intergalactic computer network". Washington, D.C.: Advanced Research Projects Agency, via KurzweilAI.net. Retrieved February 18, 2014.

[3] Delic K., Walker M., 2008. Emergence of the Academic Computing Clouds, *ACM Ubiquity Volume 9, Issue 31 (August 5 - 11, 2008)*.

[4] Fox A., 2011. Cloud Computing – What's in it for me as a Scientist? *SCIENCE, Volume 331, 28 January 2011*.

[5] Turing, A.M., 1950. Computing machinery and intelligence. Mind, 59, 433-460.

[6] Delic K., Riley J., 2013, *Current and future trends in AI,* IEEE ICAT Oct/Nov. 2013.

[7] THE ECONOMIST, 2015. *Rise of The Machines, May 9, 2015*.

[8] Farmelo G. 2014, Smashing Physics review – how the Higgs particle was found, THE GUARDIAN/Science, May 2, 2014.

[9] Turner M., 2015. Big science is hard but worth it, *SCIENCE, Volume 348, 14 April 2015*.

[10] Bigot B., 2015. Pull together for fusion*, NATURE Volume 522, 11 June 2015*.

[11] THE GUARDIAN/Science, 2015. Rosetta spacecraft to change orbit in bid to strengthen Philae communication*, June 15, 2015*.

[12] Wright A., 2014. Big Data Meets Big Science, *Communications of the ACM , Vol. 57, No. 7, July 2014*.

[13] NATURE, 2015. Maestros of grapheme, Interview, *NATURE,* Volume 522, 18 June 2015.

[14] Enserink M., 2015. *A €1 billion brain reboot, SCIENCE,* Volume 347, 27 March 2015.

[15] Venter C., 2001. *The Sequence of the Human Genome, SCIENCE,* Volume 291, 16 February 2001.

[16] Steven Weinberg, 2015. *To Explain the World: The Discovery of Modern Science*. HarperCollins Publisher

[17] Martin A. Walker, Private Communication, July 2015.

[18] Weinberg A., 2011. Impact of Large-Scale Science on the United States *SCIENCE, Volume 331, 28 January 2011*.