

An Interview with Masaru Kitsuregawa, Director General of the National Institute of Informatics

# Big Data — Leading Technology for an Affluent Future

By Japan SPOTLIGHT Editorial Section

Big Data is now a leading technology in Japan that could totally change our society. Dr. Masaru Kitsuregawa is a distinguished leading academic in this area. *Japan SPOTLIGHT* highlights his views on its future utility in human society in the following interview.

## History of Big Data

**JS:** Big Data is currently considered one of the most promising technologies to enhance our economic growth potential. First of all, I would like to ask you about the history of the development of Big Data in which you have been engaged.

**Kitsuregawa:** Before the specific term “Big Data” was created, we had been working on a “Grant-in-Aid for Specific Subject Study” organized by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). In this study, I proposed a project called “Information Explosion” in 2004 and this project ran from 2005 for five and a half years. We coined the new word “Info-plosion” in this project. This was long before the boom of “Big Data” which has now eventually become a key term in innovation, drawing much attention from the public ever since the administration of U.S. President Barack Obama started a “Big Data Initiative”.

At the beginning of our research, we needed to take note of the phenomenon whereby the quantitative flow of a wide range of information produced by modern society had increased explosively since the beginning of the 21st century. This is truly an unprecedented phenomenon in human history. At that time, we had started to discuss how to deal with this phenomenon and what would be necessary on the part of IT engineers. This was one of the largest IT research projects under MEXT for us, with around 500 researchers participating. Shortly after the end of the project, this new term “Big Data” came up in the United States. Big Data is a much simpler adjective and noun compared with Info-plosion. I think



*Masaru Kitsuregawa, Director General of the National Institute of Informatics*

Big Data is a very attractive name that people can easily understand.

Another project, organized by the Ministry of Economy Trade and Industry (METI) and called the “Information Grand Voyage Project”, had been started a year later than the one initiated by MEXT. This was a \$100 million project on Big Data initiated by METI with an industry orientation.

More recently the “Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST)” initiated by the Cabinet Office was started. From 2009, this program called for applications from the public and at first the then government of the Democratic Party of Japan called for 30 applicants from among researchers regardless of their area, providing \$90 million for each, and then after the change in power, the Liberal Democratic Party government provided \$1 billion as a fund for all the applicant academics. This was a one-time payment and this project was

finished due to budget constraints. It is very unusual for an academic researcher to have such a large sum of money for research. I participated in this program and enjoyed discussions with many competent and creative scientists. The term “Big Data” was not in use yet. However, assuming that we were heading towards a world driven by data, I set up two important research components. The first was a study on how we could deal easily with large amounts of data, as the size of data is increasing tremendously. To do this, we would need to create an extremely high-speed database engine, and with this in mind we pursued research aimed at achieving the same speed of data processing as before even though the amount of data was 1,000 times greater.

The second component was research on the application of Big

Data, in other words what could be done by using such large amounts of data with a sophisticated database engine and an analytics engine. Since it was strongly recommended that we collaborate with industries for their revitalization, we promoted these two components in collaboration with business firms.

We were given much discretion for research in this program. We could do whatever we wanted. This program was the first attempt in Japan to give researchers such large freedom in research, and we were able to attempt what nobody else had ever tried.

## Social Usefulness of Big Data

**JS: What do you think of the quality of Japanese information technology used in realizing Big Data? Could it be one of the best in the world?**

**Kitsuregawa:** It is not easy to define what would be the best in our IT domain, but I think we were successful in creating some practical outcomes with Japanese indigenous technology. For example, when you go to a hospital in Japan you pay 30% of the medical cost and the remaining 70% is covered by insurance. In this healthcare system that ensures the whole nation is covered by public health insurance, hospitals in Japan own all the patients' information with respect to receipts, namely the medical payments paid by public insurance to the hospitals. The amount of this data is so enormous — around 40 billion items each year — that it is too big to handle and we were asked to engage in collaborative research on how to deal with it efficiently. While it used to take more than a week to reassemble such big data for use in analysis, in our institute we created the data immediately for real-time analysis. This had never been achieved in Japan before.

**JS: In our aging society, management of hospitals must be as efficient as possible to meet the needs of increasing numbers of elderly patients. This technology will surely be useful for achieving efficient management of medical systems.**

**Kitsuregawa:** Yes. For example, though the Japanese Ministry of Health, Labour and Welfare believes that medical services must be spread all over Japan as equally as possible, by using our data set we can discover where there are regional differences between patients with some specific diseases, and thus medical services can be better prepared for a specific disease more frequently observed in a specific place.

More particularly, you may believe that hay fever starts in December and ends in April, but according to an analysis using our Big Data, in Tokyo there are patients with hay fever all year round. This clearly shows us that hay fever can exist regardless of pollen, which has been considered the sole cause of hay fever. Perhaps it could be said that the stressful social environment in Tokyo is another cause.

Another specific example is the case of patients with influenza. Data on influenza are available today by the month. If we can have such data by the hour, we will be able to find any schoolchild in a

specific elementary school who gets flu and goes to hospital in the morning, and then by school lunchtime a medical service team would be able to warn the pupils at that school to be well prepared for the flu by asking them to wash their hands and gargle. With such high-speed transmission of Big Data, we can change our lives dramatically.

**JS: How about other areas of life?**

**Kitsuregawa:** Another promising area for the application of Big Data technology is education. We have just started preliminary research worldwide and have not had any precise outcome yet. But I can tell you that we would be able to find very precisely how much each student understands what a teacher is saying in the classroom.

Teachers have generally been teaching unilaterally and have not been able to find out very well if their pupils have understood them. But with online e-learning systems, in particular the Massive Open Online Course (MOOC), a student can check repeatedly by replaying what a teacher is saying — and then the teacher could discover what students find difficult to understand in his or her lecture. Teacher and student could thus continue to learn from each other in a more coherent way.

This means that our education system could reduce the number of students who drop out from school because they cannot understand the lessons or lectures very well, and this would be a great improvement. It is amazing that Big Data technology could even achieve an education reform that we never dreamt about.

**JS: To strengthen security at the Tokyo Olympics and Paralympics in 2020, we could use Big Data in adopting countermeasures against terrorism and crimes, couldn't we?**

**Kitsuregawa:** Certainly. In the case of the 2012 London Olympics, they started preparations to strengthen security six years before the Games. So in 2014, six years before the Tokyo Olympics, we started discussions on a wide range of security issues, in particular how to cope with cyber crimes, which are significantly increasing today.

In the case of London, British Telecom made great efforts to find a way to prevent possible cyberattacks in stadiums, where most of the spectators could use smart phones and were thus all connected to a high-density network. In the case of the Tokyo Olympics and Paralympics, we should highlight the Paralympics as Japan is becoming an aging society more rapidly than any other country. As the percentage of elderly people to the total population increases, there will be more disabled people due to aging than before. Therefore, we have a chance to show the world and the athletes at the Paralympics how comfortably handicapped people could live and how easily they could go wherever they like in Tokyo.

I think it would be far more important for us to spend time and resources exploring how to use IT to achieve this target for 2020. Since there are many kinds of disabilities, it would be a big challenge to employ IT in the interests of such handicapped people.



## Further Innovations from Big Data

**JS: Could more detailed use of data in hospitals or schools, as you mentioned, stimulate further innovations?**

**Kitsuregawa:** Indeed. The essential merit of Big Data is to make observable the phenomena of human activities that have not been grasped so far. This means that we could see everything clearly with Big Data. It can be applied to any area, not only health care or education but also the environment, manufacturing, disaster prevention or cybersecurity, energy, transportation, and so on. Then when we can see all these social domains in detail with greater transparency, we will be able to discover the inefficient parts of our society. I think Big Data could provide us with a means to improve our sensitivity and find out what may be wrong in our society.

**JS: Renewable energy sources are expected to play an important role in our energy policy from now on and sophisticated management of energy consumption will be necessary for energy conservation and efficient utilization of these sources. This is the concept of the “smart city”. Could Big Data be applied to this “smart city” concept?**

**Kitsuregawa:** Yes. The electric power generated by renewable energy sources would be very unstable and need to be controlled to meet demand on the basis of extremely detailed data analysis. That would be a new area where Big Data could make a good contribution. In a smart city, we would need to resolve a wide range of issues including urban traffics system, crime-prevention, medical emergency systems, and measures against fire or flood, etc. In this regard, Big

Data is expected to play a key role in achieving this. Above all, Big Data must be an indispensable tool for discovering the core issues around those subjects. On the other hand, though we cannot say it would be easy to solve those problems in cities, it is true that they have large tax revenues and many business firms are concentrated in big cities so that they could put a variety of countermeasures into practice. It would be “non-city” areas like Fukushima that would face a serious difficulty in dealing with those challenges and they could be neglected by business, since their development can be considered to have a low economic impact, as in the case of curing an unusual disease hard to treat. We should not forget that what academics should truly care for are “smart non-cities” with very few inhabitants and not cities.

**JS: How about disaster prevention or earthquake prediction?**

**Kitsuregawa:** It would be extremely difficult to improve the predictability of earthquakes. However, in the case of the “smart city” concept, urban transportation systems could be improved by Big Data. For example, today we do not yet know how to deal with a huge crowd unexpectedly rushing by public transport to a stadium where a Japanese athlete may be competing in an Olympic final. But by using Big Data we could be well prepared for this situation, since we could collect data on how many people have been affected by incidents like railway accidents in the past and on the change in passenger flow between different transportation systems, and apply such data to the case of the Olympics. Railway companies could think about how many extra trains should be prepared for such an unexpected shock to the transportation network. That is how Big Data could initiate an improvement in urban city management technology, and this could be applied to disaster management as well.

## Public Policy Support for Big Data

**JS: There must be some public policies necessary to support the Big Data project. Should one of them be an accommodation of a law on authorship?**

**Kitsuregawa:** IT progress is exponential and there are great expectations over laws that could help people adapt to this progress. For example, data input in cloud computing for anybody’s use could be considered illegal in the light of authorship. However, in the US today they have adopted the concept of fair use. For example, there is a service called Google Books which, if you show some sentences from a scanned book and ask to see more of its content, responds by showing you only some parts of the book related to those sentences. This is now considered legal and fair use in the US after a long dispute.

We do not have a concept of fair use yet in Japan. Therefore, more specifically, in the case of school education, even though pupils absent from school due to flu are now allowed to “attend” classes by video link, once they have recovered and try to take the classes they missed through Big Data that I referred to as a potential tool for

initiating educational reform, this could violate the current authorship law.

Above all, Big Data is currently not covered by copyright. The key question would be who is the authentic owner of Big Data. For active use of Big Data, I believe we will need legal definitions. The evolution of IT is so rapid and significant that legal systems not limited to copyright and social acceptability must co-evolve with it.

**JS: I think we will need an alliance between academia and industry to promote the further practical application of Big Data. Would you agree?**

**Kitsuregawa:** Yes. An academia-industry alliance should be inevitable. We academics do not have any data on working factories or services in the industries. We have only data on academic research related to science and technology or social sciences. Therefore, our technology developed in universities must be applied for practical use based on the data that business firms would have. This is true not only of Big Data but also in other domains. The academia-industry partnership will become increasingly important in achieving innovation. As a matter of fact, Japanese business firms' understanding of the importance of data lags behind that of US companies. This is clear by seeing that the quantity of the storage systems that Japanese firms annually set up is much less than that of US firms.

This means that Japanese companies are not preserving data well in comparison to US firms. They do not know well exactly what data they own in many cases. Furthermore, even if they do know, there seem to be very few companies in Japan with a clear perspective on their utilization. In this light, I feel strongly that we will need to strengthen an academia-industry alliance and in the end find a framework in which a business firm could actively take advantage of various technologies developed by university institutes.

## Possibility of International Collaboration

**JS: How about international research collaboration in your area?**

**Kitsuregawa:** We have already organized a workshop for an exchange between British and Japanese researchers on Big Data in collaboration with the British Embassy in Japan. The UK people involved have visited our institutes related to Big Data as well. The French Embassy is also interested in such a workshop and we are now collaborating to arrange that. In collaboration with the US National Science Foundation (NSF), a joint project on Big Data concerning disasters is about to be started. In Asia as well, we are quite often invited to the opening ceremonies of Big Data research institutes in China or South Korea. I think such global communication to share our empirical knowledge on data will be very important in encouraging further progress in Big Data technology. Such international alliances will obviously be in Japan's interest. I guess there will be international "data driven" collaboration spreading out all over the world hereafter, making a clear distinction from the collaboration through knowledge or technology in the past.

**JS: How about countries other than the United Kingdom and France?**

**Kitsuregawa:** We have pursued a medical program in Bangladesh. You may know of Grameen Bank, which won the Nobel Peace Prize in 2006 for its contribution by microfinancing to creating social foundations for poor people. One of its services is called "village phone". This assumes that it would be difficult for everybody in a village to have a phone, so it loans a cellphone to an honest person in the village and if a call comes for any of the hundreds of inhabitants in the village, that person will tell him or her. About 10 years later, the cellphone has become widely used and the village phone has become unnecessary; so the cellphone is now greatly transforming society. This cellphone can also be used to call an emergency medical service if someone is sick. But the medical service providers cannot answer all the questions since they cannot observe the patients' symptoms directly as in the case of face-to-face consultations. So we provided such patients with a kind of "portable clinic" cellphone that can measure blood pressure, pulse and urine content, and send this data to a medical clinic in Dhaka, the capital. We did this for more than 10,000 patients, and by using such data we were able to assess what medical service system would be most desirable for the poor people in Bangladesh.

Having worked on this project for three years, we are now seeing that the local people are trying to turn our service into a business of their own. This is, I believe, one of our crucial international contributions.

## Future Projects

**JS: What will be your next topic of interest?**

**Kitsuregawa:** We have had some interesting findings. We can show you what people were talking about immediately after the great earthquake of March 11, 2011, through a record of tweets and retweets. For example, people who experienced the great earthquake in Kobe in 1995 sent a lot of practical advice to residents in the Tohoku area, such as the need to conserve water in a bathtub or to turn off the gas valves, and so on. We can now not only trace back and find data on conversations but also collect such data on tweets and retweets in real time. You can see what people around the world are most interested in and how they feel about things immediately.

We can create a sort of wish list by doing text analysis of such data with deep text analytics. For example, if we take the compound verb "run short", we can extract data to produce a list of the items which people were most worried about immediately after the disaster, such as electricity first and gas second. We can also use adjectives for analytics. We can learn what people were most afraid of then by using the same analysis for the adjective "scary" — and we get results such as further tremors and the emergency alarm noises. This is how we can now discover people's feelings very precisely.

Big Data can be used to enhance the transparency of all phenomena, including human emotions. This might be another area for further exploration.

**JS**