In this presentation, I will discuss three challenging big data applications that are thought to have promising large social benefits, but which might bring large risks at the same time. The applications are: (1) traffic data acquisition and analysis for more efficient and safer driving of vehicles on public roads; (2) traffic simulations for evacuation assistance in the event of a mega-earthquake; and (3) real-time physiological information acquisition and analysis for health care. These technologies are expected to bring many social benefits, including economic opportunities, reduction of the environmental burden, promoting health, and saving lives in disasters. On the other hand, these technologies might invade people’s privacy on a large scale and produce a serious imbalance between big organizations, such as governments and large companies, and individuals by collecting and analyzing individuals' sensitive personal data. Unfortunately, we have no easy remedies for these concerns. However, I believe we can build a good juridical and technological framework that will allow us to take advantage of the benefits of big data and minimize its disadvantages. We must have a firm grasp on these technologies and carefully analyze them to balance their advantages and disadvantages in the future.

We see big opportunities and advantages in big data applications. For example, in Japan, the public roads in urban areas are heavily trafficked because of the high population density, so there is a large social need and an opportunity to reduce traffic jams in order to make the overall traffic flow more efficient and lower the environmental load from road vehicles. Honda, a Japanese vehicle manufacturer, has provided Internavi, a smart automotive navigation system which shows traffic information and assists the human driver with real-time acquiring data and their analyzed results. This automotive navigation system gives the driver of the vehicle information about the weather at her destination, the traffic jam and car accidents on route to her destination, the most efficient route to take to her destination, and so on, making her journey safer, faster, and more comfortable than it would have been without the use of the system.

The automotive system acquires the floating car data (FCD) from each vehicle that is equipped with it every five minutes and sends the FCD to the central electric clearinghouse. The central electric clearinghouse analyzes the FCD and the data from
the VICS (Vehicle Information and Communication System), which provides road traffic information, including traffic jams and traffic controls, acquired and integrated from the police and the public road administrators around Japan. It then sends the traffic and related information, which is analyzed and integrated from the above data, to each vehicle. So, vehicles equipped with the Internavi can share and utilize the same traffic and related information to assist drivers so that they can drive efficiently and safely to their destinations, and, as a result, the overall traffic flow in the region can be more efficient and safe.

Honda expects a reduction in the fuel consumption and CO2 emissions of vehicles equipped with Internavi and, as a result, the total CO2 emissions from the whole of the traffic will decrease. Moreover, the FCD acquired from each vehicle in the five-minute intervals might be used for traffic or environmental policymaking by visualizing the data. Honda uploaded a video of the visualization of the FCD concerning CO2 emissions from vehicles travelling on the streets. Let’s watch it:

https://www.youtube.com/watch?v=L2pttzjunpw

Hence, we might be able to expect FCD or some big data collected from each vehicle or each human to be utilized for evacuation assistance in the event of a natural disaster, such as a mega-earthquake. Honda utilized FCD to support efficient and safe mobility in the disaster-struck Higashi-nihon Dai Shinsai, the Great East Japan earthquake, on March 11, 2011. Honda published the visualization of the traffic situation in east Japan which was presumed by analysis of the collected FCD. We can also see the video that Honda produced to publicize this experience on YouTube. This video lasts four and a half minutes, so it is a bit long, but, it’s interesting. Now, let’s watch it:

https://www.youtube.com/watch?v=IoIDSWBMao4

Honda made an animated video from the accumulated record of the traffic mobility visualization on the roads in East Japan over 20 days from March 11, 2011 to the end of the month, and visualized the way in which the traffic mobility and the lifeline had recovered from the mega-earthquake as time went by. Now, let’s watch the first one or two minutes of the video:

https://www.youtube.com/watch?v=36SmV91F6S4

In Japan, with the Great East Japan earthquake as a start, the government, industry and academia are struggling to gather and analyze the information which they hold internally and haven’t yet shared, in order to use the information to reduce the damage of future earthquakes and to smoothly recover from the current damage. For example, eight companies, including NHK, Google Japan, and Twitter Japan, have each gathered and analyzed earthquake-related data to understand the damage and the loss of human
lives from the Great East Japan earthquake. The analysis and the visualization of the collected mobile phone data and car navigation records can show where people were and where they moved to before, in, and after the earthquake. Hundreds of millions of tweets on Twitter about the earthquake and its damage reveal the situation of the earthquake-stricken area in considerable detail. The analysis of collected tweets about the crops and the fish around the earthquake-stricken area, especially near the atomic power plants destroyed by the earthquake, Fukushima Dai-ichi Genpatsu, for one and a half years make it clear that the harmful rumors on the Internet became a large obstacle, preventing the progress of the recovery of the agricultural and fishery industries in the area. The big data, which suggest a change in the pattern of the transactions among Japan’s companies and of the pattern of employment, showed that the behaviors of the companies outside the earthquake-stricken area might have led to the delayed recovery from the damage in the area.

Thus, we have several projects which may contribute to reducing the damage and the loss of human lives in future mega-earthquakes using big data applications. We can see hope for big data applications.

In addition, we may have another hope of using big data in health and medical applications, which would continuously acquire physiological data from persons in everyday life by high time density. The development of small and lightweight wearable health care measuring devices with various sensors and microcomputers has enabled the easy and constant collection of physiological data from the bodies of persons wearing the devices.

For example, a mobile phone or smart phone with an accelerometer can constantly acquire and record data that suggest the movement and/or posture of the person wearing the device; for example, some applications count the steps a person takes while wearing the phone and calculate the time and distance walked or ran.

A few years ago, many human activity trackers, bracelet-type devices used exclusively for tracking and recording the activity of the person wearing it in daily life, started to be sold. These devices do not only tell the time when the user walked or ran, but also when he or she slept from the data acquired from the acceleration sensor. They generally have Bluetooth communication features and communicate with mobile phones or smart phones to send and record the data acquired from the person wearing it. So, the user can see her activity record on the phone’s display and analyze it to improve her life pattern and her health. Recently, wearable pulse monitors with built-in microcomputers and accelerometers have become available, which can acquire and record more detailed data about a user’s activity. The manufacturer also provides a professional service that
advises the user about health care based on the analyzed data acquired by the tracker and accumulated from the user’s daily life.

In the future, the development of electronic sensors and microcomputers will provide wearing devices with more sensors which will constantly acquire and record various physiological data in our everyday lives and in heavy time densities. These will help us improve and maintain our health. We are hopeful about the potential of big data from such devices.

Several global electronics manufacturers and system integrators, including Japanese companies, are trying to develop a new market for wearable health care devices, and many videos predict a bright future for these. Let’s watch one now produced by NTT Data to advertise their technological future:

https://www.youtube.com/watch?v=AoP8PbGX5g0

However, a serious concern with the good news of the big data applications shown above is that they deal with sensitive data, such as tracking a person’s movements and activities, which may reveal some private facts and health-related data, including telling the person about a serious disease.

We only need statistically processed data of people’s movements and activities to understand the overall tendency of traffic and realize its efficient and safe flow for the whole region. The raw individual tracking record of each vehicle or each person is acquired and integrated to analyze the whole tendency, but it is not needed directly for this purpose. So, we usually have no risk of privacy invasion through wrongdoing of the individual tracking record if the company acquiring the data keeps its contract with the users. Typically, such contracts state that the raw data will not be abused or misused nor will it be provided to third parties, and that the confidentiality and the integrity of the data from accidents and attacks from outside or inside will be protected.

Because the Act of the Protection of Personal Information in Japan enacted in 2003 does not protect the tracking data of vehicles or a person’s personal information, a company that acquires such data can use it and sell it to third parties without any constraints and limitations of the law, as long as the company removes the direct identifiers of the data subjects from each record within the data. In the other words, the data derived from a person is not regarded as personal information in the legal framework if direct identifiers of the data subject are removed.

However, even if the identifiers of the data subjects are removed, we can determine quite a lot by combining the data subject and its attributes through big data analysis because of the enormous developments in computer technology. So, some jurists, ICT scientists, and policymakers are concerned about big data privacy, even if the data is
modified to remove the identifiers, and have tried to change the definition of personal information in Japan’s Act of the Protection of Personal Information. In 2014, the policy to revise the act was almost completed and it is planned to be installed as a revised act this year.

Unfortunately, information scientists have found that we cannot completely anonymize personal data, so we will always have some risk of privacy invasion when utilizing personal data, even after modifications to anonymize it. On the other hand, the statistical data gathered from a combination of personal data is at low risk of invading the privacy of the data subjects, since the statistically processed data cannot generally reveal details of the situation at any significant level of granularity. Thus, we have a tradeoff between privacy protection and detailed data analysis.

Furthermore, we also have some privacy invasion risks concerning health care data. Health care data doesn’t look sensitive because it is not a medical record as it stands. However, if the health care data, which are recorded constantly and in heavy time density, were analyzed properly, they might reveal a medical situation of the data subject in detail. So, detailed health care data from wearable devices should be protected for confidentiality in the future. Now, I heard that some ISO processes are making progress toward implementing standards for the protection of health care data. We need to regard such data as sensitive and properly protect it through technology and the legal system.

Efforts to establish effective measures to protect our privacy and reduce the risks are under way. I believe we can build a good juridical and technological framework that will allow us to take advantage of the benefits of big data and minimize its disadvantages. We must have a firm grasp on these technologies and analyze them carefully to balance their advantages and disadvantages in the future.